
SMITHSONIAN INSTITUTION LIBRARIES

THE NAUTILUS

Supplement 2
(Issued with Volume 108)
ISSN 0028-1344

Molecular Techniques and Molluscan Phylogeny

Proceedings of a Symposium held at the
Eleventh International Malacological Congress
Siena, Italy

31 August-5 September, 1992

Edited by

M. G. Harasewych
Department of Invertebrate Zoology
National Museum of Natural History
Smithsonian Institution
Washington, DC 20560 USA

and

Simon Tillier
Laboratoire de Malacologie
Muséum national d'Histoire naturelle
55, rue Buffon
F-75005 Paris, FRANCE

1994

Available from:

THE NAUTILUS
P.O. Box 7279
Silver Spring, MD 20907 USA

THE NAUTILUS

Supplement 2
(Issued with Volume 108)
August 11, 1994
ISSN 0028-1344

CONTENTS

Preface

With its tremendous diversity and excellent fossil record extending from the earliest Cambrian to the present, the phylum Mollusca presents extraordinary opportunities and challenges to students of all aspects of evolutionary biology. The confluence of new techniques of data acquisition (among them scanning and transmission electron microscopy and nucleic acid sequencing) and data analysis (most notably cladistic methodology) has prompted many advances in studies of molluscan phylogeny at all levels over the past decade. Yet despite the use of varied data sets and techniques, many of the most basic questions at the highest taxonomic ranks remain unresolved. While mollusks, particularly *Cepaea*, *Cerion*, and *Partula*, have played important roles in the studies of population genetics and speciation, the phylum remains underrepresented in studies at higher taxonomic levels. Happily, this situation is changing, and increased research in molecular malacology is evidenced by the appearance of the "Mollusc Molecular News" and bulletin board on the internet.

The papers comprising this volume were presented at the Symposium on Molecular Techniques and Molluscan Phylogeny that was convened during the Eleventh International Malacological Congress, held in Siena, Italy, in August–September of 1992. This collection of works contains the majority of the papers and posters presented during this Symposium, and represents a overview of contemporary research in this rapidly growing area of systematic malacology. Although a symposium on a similar topic was held during the American Malacological Union meetings in 1987, the results presented at that time were judged by the participants to be too preliminary to warrant publication. We are heartened that the majority of the contributors to the present symposium felt that their work was sufficiently advanced to be published.

We thank Prof. Folco Giusti for the invitation to organize this symposium, and for his assistance in coping with the many attendant details. We extend our appreciation to all participants in the Symposium, and recognize the special efforts that several colleagues have made to attend these meetings. The contributions of the numerous co-authors, many of whom were not present at the Symposium but nevertheless contributed to its success, are gratefully acknowledged.

In the Plenary Lecture, which opened the Congress and set the stage for the Symposium, George Davis set forth his views on the utility of molecular genetics for molluscan systematics. Presenting case studies primarily from his own work, he documented the value of molecular data as well as the potential pitfalls in incautious interpretation of molecular data.

The organization of the Symposium into three sessions, each covering a broad topic, is reflected in the arrangement of papers in this volume. The first group of papers documents the use of population genetics techniques for studies of speciation or phylogenetic inference. Colgan and Ponder present a sophisticated analysis of the varied effects of restrictions on gene flow on speciation using allozyme data. Emberton reviews and compares methods for constructing phylogenies using allozyme data, while Adamkewicz and Harasewych explore the utility of RAPD techniques, recently developed for differentiating populations and strains, for the inference of phylogenetic relationships among closely related species.

The second group of papers focuses on the mitochondrial genome and explores the utility of mitochondrial DNA (mtDNA) for investigating a broad range of genetic relationships. Boore and Brown review the varying and contradictory hypotheses regarding the relationships of the phylum Mollusca and its classes, and advocate the use of the arrangement of genes (gene order) of the mitochondrial genome as phylogenetic characters. Terrett and colleagues present new data on the gene order of *Cepaea nemoralis* and compare it with those of other metazoans. At the other extreme, Cook and Zouros analyze inheritance patterns of variations in the size of the mitochondrial genome among sibling scallops, and conclude that because of its rapid turnover, such size variation does not provide information useful for taxonomic studies. In the last paper of this group, Rumbak and colleagues use the sequence of a portion of the gene for the small ribosomal RNA to study the relationships of 13 species of littorinids.

The third group contains papers on the use of ribosomal sequence data to resolve phylogenetic relationships among mollusks. Winnepeninckx and associates present the complete sequence and structure of the 18S rRNA of *Onchidella celtica*, a pulmonate snail, and compare it to 25 other known metazoan sequences to assess the monophyly and relationships of three molluscan classes as well as the relationships of Mollusca among the Metazoa. Preliminary analyses of sequences derived from the D6 loop of the 28S rRNA of 43 gastropod and bivalve species reported by Rosenberg and collaborators reveal variation in the rates of sequence divergence but do not refute morphology-based phylogenies. A group led by S. Tillier present the most broadly represented sequence-

based phylogeny of Gastropoda to date. Geller and Powers use site-directed mutagenesis to descriminate between sibling species of *Mytilus* based on a single base difference in a region of their 16S ribosomal gene.

Two additional papers, originally presented as posters, are included. One (Lazaridou-Dimitriadou and colleagues) reports on allozyme variation in Greek populations of *Helix aspersa*, the other (Backeljau and numerous collaborators) reviews the use of isoelectric focusing in molluscan systematics.

Initial attempts to investigate the origins and early evolution of the Mollusca by the use of molecular data have met with only limited success, and in the process have questioned some of the most basic precepts of morphology-based classification. Perhaps most notable of these is the growing body of evidence from both nuclear and mitochondrial genomes that places Bivalvia as the outgroup to a clade containing the Gastropoda and Polyplacophora. Clearly, the addition of taxa, especially from presently unrepresented molluscan classes, would be of great value.

The growing body of molecular data, together with increasingly sophisticated and better reasoned methods of analysis, portend great advances in our understanding of molluscan evolution in the coming years.

M. G. Harasewych
Simon Tillier

Molecular Genetics and Taxonomic Discrimination

George M. Davis
Pilsbry Chair of Malacology
The Academy of Natural Sciences
1900 Benjamin Franklin Parkway
Philadelphia, PA 19103, USA

PROLOGUE

The following paper is based on the Plenary address I gave before the 11th UNITAS Congress, the International Congress of Malacologists, held in Siena, Italy 28 August–7 September 1992. The address is dedicated to Professor Folco Giusti, a friend, colleague, and scholar dedicated to excellence in systematics who, with G. Manganelli (1992), wrote concerning the discrimination of species:

"The good systematist is not one without doubts or the one who always succeeds in defining a phenomenon or recognizing a "species". It is rather the one who studies the phenomenon trying to understand it in all its facets and who is not above admitting that its exact nature escapes him . . .Let us stress again that only a little humility and a little consciousness are required."

INTRODUCTION

If we could call back to the present some of the early fathers of morphology-based malacology, for example Cuvier, Bouvier, Troschel, Stimpson, Pelseneer, Thiele, Johansson, and Pilsbry to name a few, and bring them up to speed on the vast accumulation of literature since their time, they would readily understand and be enthusiastic about the modern day potential for sophistication in taxonomic discrimination. They would say that it is about the recognition of, and the definition of species, genera, and higher taxa. They would be in agreement, and I with them, that the fundamental basis for taxonomic discrimination was then, and is today, the comparative anatomical data set.

Unfortunately, over the past several decades, detailed comparative anatomy has been the most under-used tool in molluscan systematics. However, given the recent acceptance of cladistic methodologies (in malacology, only in the past 6 to 7 years), one sees a return to comparative anatomy. The hunt for unique anatomical characters and character-states in order to nest taxa in sets based on synapomorphies, is gaining increased respectability. Growing awareness that there is a need for well defined qualitative anatomical characters and their states, which serve to demonstrate differences among taxa in order to construct hypotheses of evolved relationships (phylogenies), should stimulate modern anatomical work on all groups of mollusks.

Molecular techniques have long been used as an aid for discriminating among taxa. However, as with cladistic tools, malacologists have lagged far behind microbiologists, mammalogists, and herpetologists in applying them. The use of immunology in systematics is over four decades old; the use of allozymes, three decades. The now-generation is scrambling to sequence RNA and DNA aided by PCR and cloning.

The use of allozymes in molluscan systematics is now well established and will not be supplanted by genomic techniques for years to come. The quantities of useful information that can be gained through allozyme electrophoresis are enormous and can be obtained at relatively little cost compared to the considerable expenses involved in pursuing sequence work. Allozymes are especially useful in comparing closely related genera, species within a genus, and sorting out species-level problems. Allozyme electrophoresis is an ideal tool for population genetics as applied to delineating species. DNA-RNA sequencing is in its infancy, literally exploding in dimensions of use, problems, and surprises. Given 600 million years of spectacular molluscan evolution, one can be sure that molluscan DNA from more than 100,000 living species from seven classes will yield numerous surprises and cause researchers to have many a migraine headache.

A major concern with molecular data is the analysis of data. It is generally agreed today that there is as yet no truly satisfactory way to analyze such data to adequately portray relationships (exhaustively reviewed by Buth, 1984 for allozymes and by Swofford & Olsen, 1990 for sequence data). There is a dichotomy of approach: cladistic and phenetic. In a phenetic mode, the usual approach is a UPGMA treatment of distance data based either on allele frequencies (allozymes) or sequence differences. A phenogram is the standard presentation. Cladistic analysis requires a unit character. With allozymes, using the locus as a character seems to be the best at present, with different allele combinations scored as character-states. With sequence data, the unit character sug-

gested is the gene. Further, genes must be calibrated for the taxa studied relative to rates of evolution if credible phylogenies are to be structured. Phylogenetic hypotheses are based on cladograms derived from computer programs such as PAUP (Swofford, 1983) or HENNIG–86 (Farris, 1988) that use parsimony criteria to obtain the shortest possible tree. However, there is a universal call by those involved in the evolution of bacteria, viruses, and protists as well as other groups of taxa for research to provide better modes of analyses of molecular data to make better trees (Davis, 1994).

One rarely sees both phenetic and cladistic analyses together in the same paper, especially when the database is an anatomical one. I strongly urge that both be used and the results examined for congruence. In the phenetic mode, combining principal component analysis (PCA) and multidimensional scaling (MDS) with ordination diagrams usually yields considerably better results than simply using similarity or distance coefficients to do a UPGMA structured phenogram. The benefits are:

1) As one moves from the initial phenogram to PCA to MDS, a tree of relationships between each taxon (species, individual, etc.) such as a Prim Network, usually becomes shorter and the cophenetic correlation with the original matrix increases indicating that the result better portrays relationships among the taxa (phenetic parsimony, Davis *et al.*, 1994a).

2) Ordination diagrams of individuals or species or genera in 1×2, 1×3, 2×3 dimensional space with taxa connected by a Prim Network provides a considerably improved understanding of taxon interrelationships both as to scale of divergence and direction of divergence of the taxa. Ordination of taxa in n-dimensional space removes the constraints of the one dimensional phenogram.

3) The PCA allows one to assess character correlations that are the basis for the distribution of taxa along each dimension of n-dimensional space. The utility of these multivariate techniques will be highlighted later in this paper.

It has been a decade and a half since I reviewed experimental methods in molluscan taxonomy (Davis, 1979b). While it was possible then to review nearly the entire literature pertaining to molluscan molecular systematics from amino acid work, through immunology and protein electrophoresis, it would be counter-productive to attempt to do so now. Instead, I will give my views on the utility of molecular genetics today for molluscan systematics. In outline form below are presented the topics I will discuss as to the prominent uses of molecular genetics:

I: To uncover cryptic species.
II: For population genetics:
  A: To detect and study hybridization.
  B: To examine special selective pressures.
  C: To study breeding structure.
III: To study patterns of evolution:
  A: Speciation.
  B: Phylogeny.
IV: To uncover unique aspects in the evolutionary process.

In presenting case studies, primarily from my own experiences with these issues, there are several key and central issues to focus on and always keep in mind:

1: There is no universal molecular clock, and taxa therefore must be calibrated relative to genetic distance. Different clades may have evolved at different rates.

2: Genetic distances do not define species or higher taxa.

3: Species concepts are important relative to interpreting genetic distance data.

4: Definitions of taxa and discrimination among taxa fundamentally require anatomical ground-plan data including developmental and cytological data, not genetic distances.

5: When morphological data yield few characters and character-states that enable discrimination among taxa, the need for molecular genetic data increases.

CLOCKS AND CALIBRATION

It is not the purpose of this paper to exhaustively review all that has been written on this topic. It is useful to abstract five key points derived from the reviews of Templeton (1980, 1981), Barton (1989), Harrison (1991) and the diverse publications of Gillespie (1984, 1986a,b; 1987) that focus on the issues of molecular clocks. 1) Rates of molecular evolution are more variable than expectations based on a simple Poisson mutation process. 2) Protein - DNA evolution rates are not consistent with neutral theory. 3) Molecular evolution is rate-variable and episodic, and, in the view of some, best explained by invoking natural selection (Nevo & Beiles, 1988; Skibinski & Ward, 1982; Murray *et al.* 1991). 4) Any attempt to apply a molecular clock to comparisons of even closely related species is hazardous. 5) There is no simple pattern between mode of speciation and genetic distance.

With due respect, one must consider the serious arguments of those supporting neutral mutation theory (refer to numerous papers by Nei and Kimura reviewed in Nei & Graur,1984) employing statistical tests of data derived from allozymic data providing heterozygosity or gene diversity where effective population size and mutation rates are known. But there is so much more involved in considering rates of evolution; one example is provided. Consider the rapid duplication and loss of genes coding for the alpha chains of hemoglobin where differences in rate are apparently associated with differences in lengths of non-coding regions (Zimmer *et al.*, 1980). They consider that adaptive evolution may depend more on the type of genetic variability than on various point mutations that affect protein structure.

That there is no universal clock, and that different clades may evolve at different rates based on molecular genetic data is clearly demonstrated in Figure 1, where regressions 1, 3, and 6 pertain to different groups of teleost fish (from Hillis & Moritz,1990 based on data from Avise & Aquardo, 1982). For further elucidation

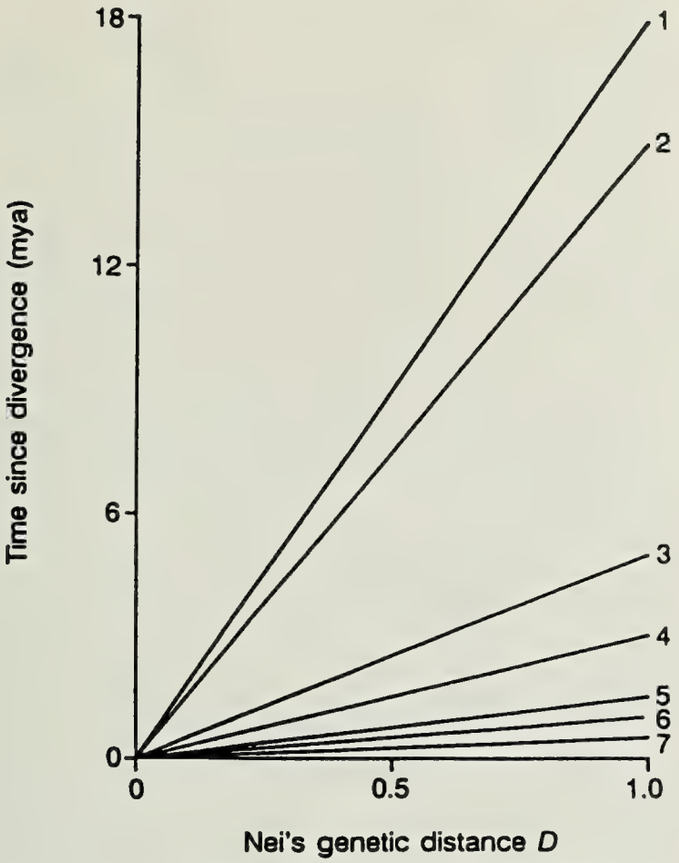


Figure 1. Regressions of changes of Nei's genetic distance D through time for different clades. Clades 1,3,6 are for different teleost fishes. Redrawn from Hillis and Moritz (1990).

on this topic, refer to Hillis and Moritz (1990) for similar plots for DNA, RNA, etc. To provide an example, a statistical analysis of DNA sequences from artiodactyls, rodents and primates show that there is no global molecular clock in mammals. Rates of nucleotide substitutions in rodents are approximately four to eight times higher than in higher primates(Li *et al.*, 1987) Regressions for different molluscan clades are similar to those shown in Figure 1, as demonstrated below.

Genetic Distances, Their Utility and Calibration:

Different measures of genetic distance have different utilities. When referring to genetic distance based on allozyme "alleles", one must state whose formula was used. The traditional D is that of Nei (1972) although the Nei (1978) modification is now widely used; it represents the accumulated number of codon substitutions per locus since time of divergence; it is a squared distance that rises with time. There is no upper limit, constraints being only the number of loci used and alleles found. It is non-metric. One of the beauties of this measure is that as taxa diverge more and more, the difference is not squeezed between 0 and 100%. Yes, there are genetic interpretation and statistical problems with values exceeding 1.0, but aside from the problems of metricity, the cause for the rise above 1.0 is clear from the data.

Sewell Wright (1978) considered Arc distance (Cavalli-

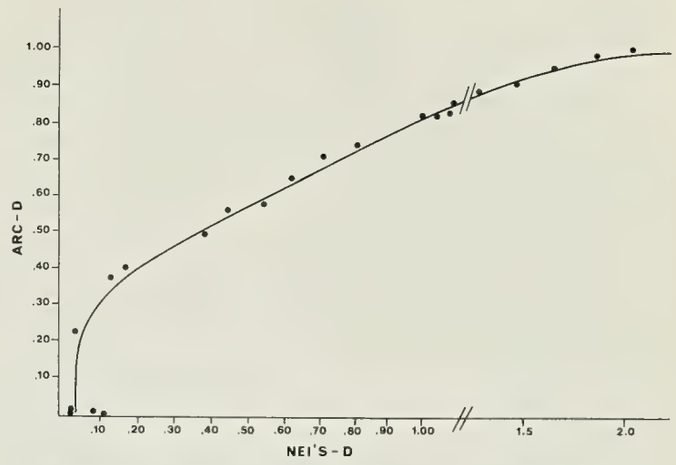


Figure 2. A comparison of Nei's D with Arc distance based on data from my various papers. As Arc distance becomes compacted, Nei's D continues to rise. As Nei's D is compacted at lower values, Arc distance gives a better approximation of differentiation among closely related taxa.

Sforza & Edwards, 1967) to be the best as it is a metric distance where the "coordinates in the hyperspace are the square root of the allele frequencies, a procedure that locates all populations on the surface of a hyper sphere with respect to a locus". Best or not in the pure mathematical sense, both distances have their use. A plot of Nei *vs.* Arc D values is given in Figure 2 based on a variety of data I have published or accumulated. It is clear that Nei's D values are compacted at the lower values (to the left) where closely related populations are compared. In this situation, Arc D gives a better understanding of differentiation. At the other end of the plot, with the greater divergence of distantly related taxa, the Arc D values become compacted and less informative while Nei's D values continue to rise thus providing a better understanding of divergence even though the values do not rise linearly and are not metric. Nei's D can be used to calibrate within-group variance, and gaps between nested sets of taxa that represent different taxonomic levels. Nei's D can be used for multidimensional scaling and with a Prim network to obtain a three dimensional visual sculpture of relationships, patterns of divergence, and gaps between sets of taxa as long as one understands the constraints and limitations due to the non-metricity of the measure. When publishing results, I advocate providing both Nei's and Arc distances. Certainly the Nei's D is needed because of the vast literature now built up using this measure. It has become a universal standard for comparison.

Beware of those who grind-run gels-publish without studying the patterns and processes of morphological diversification, life history diversification and ecological diversification throughout the clade of concern and in sister clades. There is so much more involved in understanding species differentiation then examining a genetic distance and extrapolating taxonomic rank. This is well understood in vertebrate literature; see the review of Avise and Aquardo (1982). However, I have heard it

Table 1. Calibrating clades for Nei's D: selected examples. Mean ± standard deviation.

	Population level	
Hydrobiidae (Davis *et al.*, 1989)		
6 populations	*Hydrobia truncata*	0.008 ± 0.005
Truncatellidae (Rosenberg, 1989)		
4 populations	*Truncatella pulchella*	0.029 ± 0.020
2 populations	*Truncatella caribaeensis*	0.067
Pomatiopsidae (Woodruff *et al.*, 1988)		
7 populations	*Oncomelania hupensis quadrasi*	0.039 ± 0.038
Planorbidae (Mulvey *et al.*, 1988)		
6 populations	*Biomphalaria glabrata*	0.103 ± 0.068
Planorbidae (Bandoni *et al.*, 1990)		
12 populations	*Biomphalaria pfeifferi*	0.053 ± 0.066
Unionidae (Davis *et al.*, 1981)		
5 populations	*Anodonta cataracta*	0.034 ± 0.038
11 populations	*Elliptio complanata*	0.065 ± 0.039
	Species: differences among species of a genus	
Truncatellidae (Rosenberg, 1989)		
5 species	*Truncatella*	2.226 ± 0.862
Succineidae (Hoagland & Davis, 1987)		
2 species	*Novisuccinea*	0.004
2 species	*Oxyloma*	0.269
Unionidae		
7 species (Davis, 1981)	*Elliptio*	0.210 ± 0.017
3 species (Davis, 1983)	*Uniomerus*	0.308 ± 0.165
3 species (Kat, 1983a)	*Anodonata*	0.457 ± 0.073
6 species (Kat, 1983b)	*Lampsilis*	0.609 ± 0.478
Sphaeriidae (Hornbach, 1980)		
4 species	*Sphaerium*	0.568 ± 0.310
	Genera: differences among genera of the same tribe	
Unionidae (Davis, 1981; Davis *et al.*, 1981)		
4 genera	Amblemini	0.651 ± 0.275
3 genera	Pleurobemini	0.243 ± 0.086
	Subfamilies: differences among subfamilies of Unionidae (Davis, 1981; Davis *et al.*, 1981)	
3 subfamilies		1.903 ± 0.186

remarked by some malacologists newly come to the use of allozymes that the average Nei genetic distance(D) among populations of a species is ≤ 0.09(or some such value), or that D ≥ 0.20 or 0.30 indicates a species difference, especially if the populations involved are separated by considerable distance (widely allopatric). Woodruff *et al.* (1988) have done this while applying an evolutionary species concept (e.g. Wiley, 1981) [or one could equally use a phylogenetic species concept (for a review, see Cracraft, 1989)] to argue that populations with a D of 0.60 are distinct species as they have their own evolutionary history and can be diagnosed on unique qualitative characters (e.g. unique alleles). Are they correct? No, not on this evidence alone. There are cases of perfectly good species with D ≤ 0.06 (Sene & Carson, 1977; Kirkpatrick & Selander, 1979; Davis *et al.*, 1981) and cases with populations with a large D value not being considered different species. Consider the complexity of assessing taxonomic meaning of D for mole rats of the superspecies complex of *Spalax ehrenbergi*, in Israel, where members differ by a D of 0.039 compared with a D of 0.234 between the two superspecies *S. leucodon* and *S. ehrenbergi*, indicating the recency of divergence in the later and considerable age of divergence from the former (Nevo, 1991). These issues will be revisited in this paper.

Nei's D is used to calibrate clades because one can compare among taxa from populations to subfamilies or families for the very reason that there is no upper limit for Nei's D. In this context, refer to Table 1. At the population level, it is instructive to know that allopatric populations of *Hydrobia truncata* (Hydrobiidae), spread out along the coast of North America from New England to Maryland have a mean D of 0.008 while in species of

a sister family (Truncatellidae), populations of the same species distributed from Florida through the West Indies differ by a mean D of 0.029 or more, i.e. four times as much difference as in the former case. But, in the pulmonate Planorbidae, widely allopatric populations of *Biomphalaria glabrata* differ by a mean D of 0.103 (Mulvey *et al.*, 1988)! Studies of 12 populations of *Biomphalaria pfeifferi* widely distributed throughout Kenya revealed that two pairwise comparisons had Nei's D > 0.240 (Bandoni *et al.*, 1992) while the average D was 0.053.

Genetic distances can point out anomalies, but do not explain them. Three uses in Table 1 illustrate: 1) the D among five species of "*Truncatella*" is 2.226, considerably more than between three subfamilies of Unionidae (1.903). The data suggest that more than one genus is involved. Is this correct? One needs to go back to the detailed comparative anatomy to look for marked changes in groundplan among species groups. Rosenberg (personal communication) showed me that his anatomical data would indeed justify different generic rank given the types of differences that do occur among the taxa he studied relative to the types of differences that justify generic discrimination among taxa of sister rissoacean families (Hydrobiidae, Pomatiopsidae, etc.). He awaits data for more species before naming genera. 2). In the pulmonate family Succineidae, in one instance perfectly distinct species differ by a D of 0.004, while in another the difference is 0.269. The former should be populations according to some who follow a D formula to assign taxonomic rank, while the latter data do suggest different species. I will revisit this case later. 3) In the freshwater bivalve *Elliptio*, the difference among species is low, 0.210 while in another clade (different tribe) the difference among *Lampsilis* species is three times greater, i.e. 0.609.

At higher taxonomic levels, it is clear that differences seen among subfamilies of Unionidae that go back at least to the Cretaceous accumulated slowly while much more recently evolved "genera" of Truncatellidae (given the ages of the West Indian islands for these terrestrial taxa) differentiated genetically at a considerably greater rate.

THE UTILITY OF MOLECULAR GENETICS

I: Cryptic Species and Convergence

The following example abstracted from Davis (1983) serves four purposes: 1) It demonstrates the power of molecular genetics to uncover cryptic species. 2) It clearly demonstrates nature's perversity in obscuring genetic diversity under the cover of convergent morphologies. 3) It shows that when too few morphological character-states serve to distinguish generic groupings of species, molecular genetic data may be decisive. 4) It underscores the fact that shells provide the least valuable data for elucidating relationships among higher taxa (genera, tribes, etc.).

The North American freshwater clam genus *Uniomerus* was long considered to have only one species, *U. tetralasmus* (Johnson, 1970). The genus is barely discernible from *Elliptio*. The shells of both genera are highly convergent for most species (differing only in beak sculpture that is mostly eroded away at an early stage of growth), and the only anatomical discriminant found thus far involves the complexity, or lack of it, in the branching of the branchial papillae.

William Heard and I collected what appeared to be a single population of *Uniomerus* from an area of some $100m^2$ in a stream of the panhandle of Florida. Electrophoretic results (14 loci, 24 alleles) revealed two intermixed but distinctly different species (the shells were numbered, thus allowing separation of shells based on results of unique alleles). After the fact, I could separate the species on the basis of shell for about half of the shells. I subsequently obtained a population of *Uniomerus* from Georgia with a distinctive shell phenotype and did a three-way electrophoretic comparison. To my considerable surprise, the greatest D in the three pairwise comparisons was 0.498 (from the two Floridian taxa), a value greater than between any of the 13 pairwise comparisons for species of non-lanceolate *Elliptio* (greatest D of 0.446; Davis *et al.*, 1981; Davis, 1984). Three species of *Uniomerus* were involved, not one! Parallelisms in shell characters disguised the two sympatric Floridian species. The myth that ecologically induced variation in shell size and other shell characters was generally considerable within unionid species, and thus there was only one species of *Uniomerus,* was exploded.

The question arose: Could one clearly demonstrate the relative value of shell data, morphological data, and molecular genetics data for discriminating evolved relationships among species of Unionidae? To answer this question, I compared the following taxa: The three species of *Uniomerus*, *Elliptio complanata*, *Fusconaia flava*, *Lampsilis teres*, *Quadrula quadrula*, *Quincuncina infucata*. These taxa were classified by Davis and Fuller (1981) into three tribes on the basis of comparative immunology and anatomy as follows: Pleurobemini: *Elliptio*, *Fusconaia*, *Uniomerus*; Amblemini: *Quadrula*, *Quincuncina*; Lampsilini: *Lampsilis*. Too few anatomical characters have been found to allow a definitive cladistic analysis for unionid genera. Davis and Fuller (1981) found eight, of which two were shell characters. The larger clades (tribe, subfamily levels) are readily separated, but not the differences among some genera such as *Uniomerus* vs. *Elliptio*. Thus, anatomical data are limited.

I compared the above taxa using shell morphometric data and molecular genetics. I used the same multivariate procedures with both data sets, i.e. multidimensional scaling (MDS) with ordination of taxa on the first two principal components. The taxa were then connected with a Prim network. In the morphometric analysis, 39 characters were scored and treated by Principal Component Analysis to assess character correlations prior to MDS. Results are shown in Figures 3 and 4. Results based

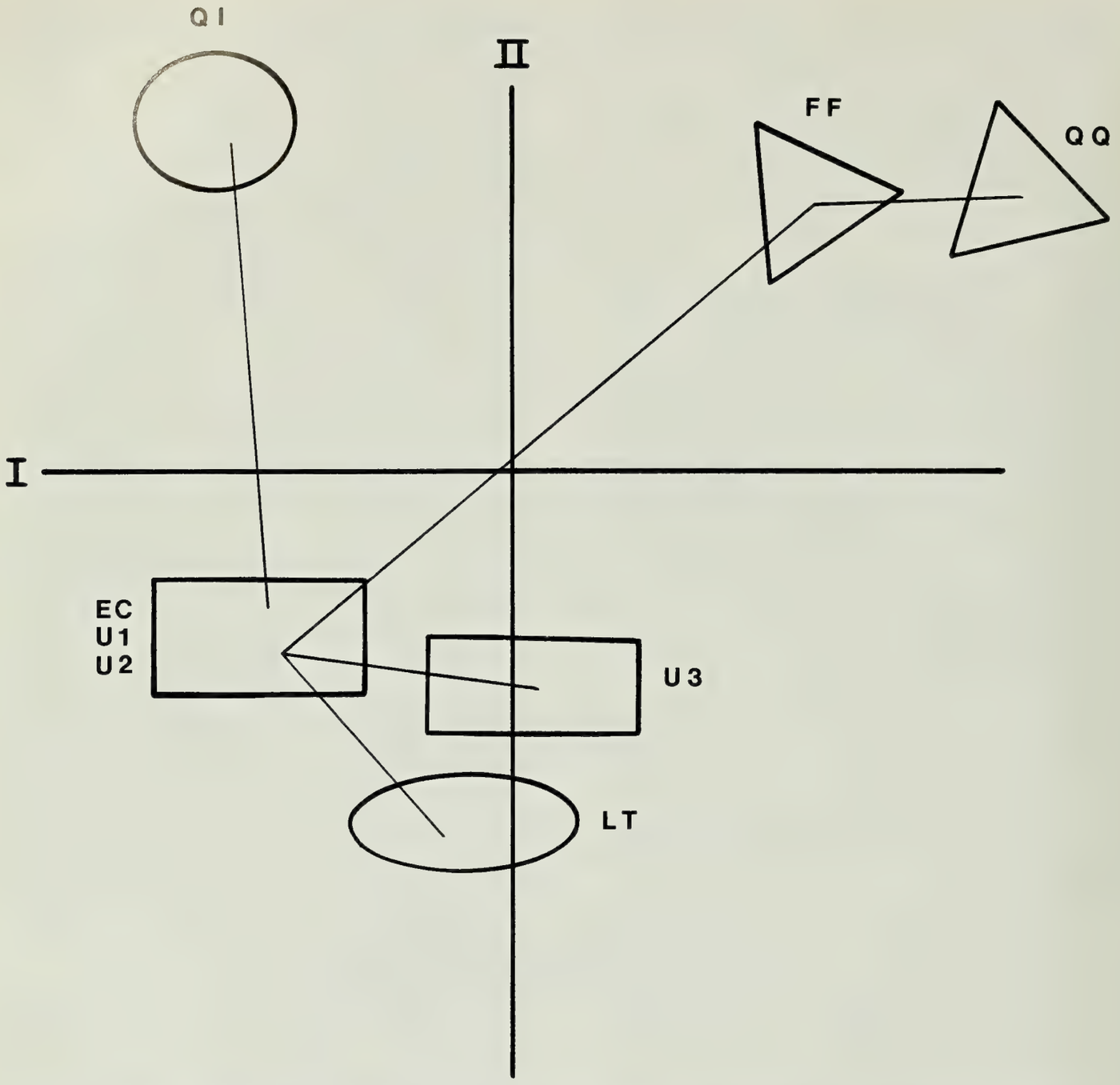


Figure 3. Ordination of unionid taxa on first two principal components following multidimensional scaling and use of Prim network; based on shell morphometric data. The shapes approximate the shapes of the taxa. Three taxa could not be discriminated (EC,U1,U2). EC, *Elliptio complanata*; FF,*Fusconaia flava*; LT, *Lampsilis teres*; QI, *Quincuncina infucata*, QQ, *Quadrula quadrula*; U1,2,3, *Uniomerus species 1,2,3*. Adapted from Davis (1983).

on shell data indicate relationships based on correlates of shell shape, not those based on anatomy and immunology. One cannot separate individuals of *Elliptio complanata* and two species of *Uniomerus*. The data based on allozymes (Figure 4) reflect the tribal relationships based on anatomy and immunology. Clearly there are shell shape convergences (Figure 3). In Figure 4, *Fusconaia flava* and *Elliptio complanata* group close together in the same computer calculated set (D = 0.208) with a D less than that between the latter species and *Uniomerus* species 1 (D = 0.303), in spite of the fact that the shape of *Fusconaia* is the same as that of *Quadrula* and distinctly different from the shape of *Elliptio* and *Uniomerus*. *Lampsilis* is genetically highly divergent from the other taxa as it should be, based on anatomy and immunology. The advantages of using MDS and the Prim network are clear. One can see the pattern of divergence of species of *Uniomerus* from species of *Elliptio*. One can see the direction of divergence of species from each other. It is readily observed that the greatest

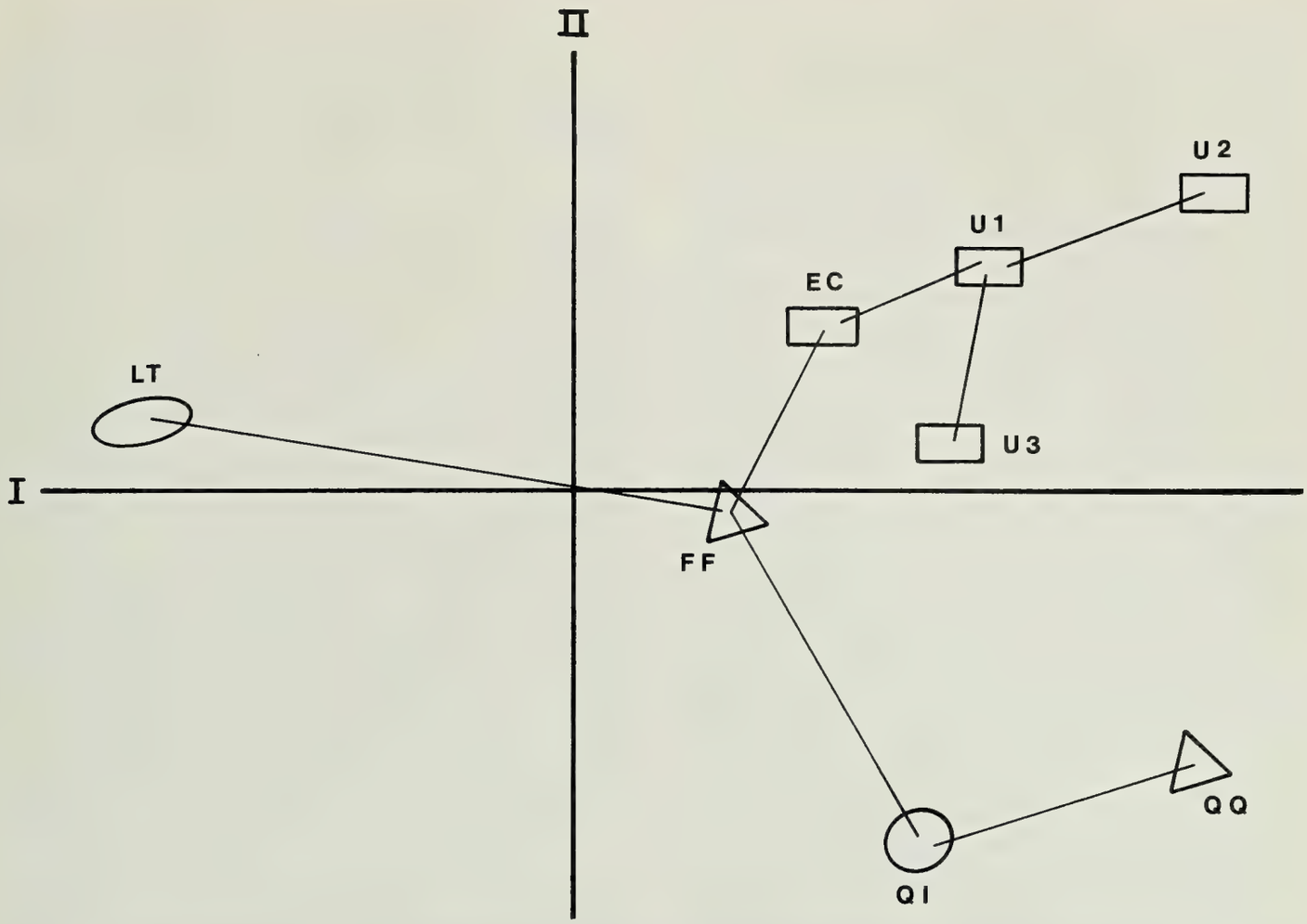


Figure 4. Same taxa and ordination techniques as in Figure 3 but using genetic distances from allozyme electrophoresis. All taxa are clearly separated. *Fusconaia* and *Quadrula* are not closely allied as in Figure 3. *Fusconaia* is closely allied genetically to *Elliptio* (contrast Figure 3). There is greater distance between U2 and U3 than there is between *Elliptio* and *Fusconaia*! Adapted from Davis (1983).

distance between two species of *Uniomerus* exceeds the distance between two genera, i.e. *Elliptio complanata* and *Fusconaia flava*; however, the direction of divergence of *Fusconaia* is away from *Uniomerus* relative to *Elliptio*.

Take Away Messages

Several points can be made:

1) There is a cryptic radiation of species of *Uniomerus* hidden by parallel evolution of shell shape and by conchological variation that was previously interpreted as ecophenotypic or random variation.

2) It is predicted that other species of *Uniomerus* will be found. Consider the type species, *U. tetralasmus* from Texas that has a shell phenotype different from the phenotypes of the three species discussed here. From the evidence thus far, populations of *Uniomerus* with different shell phenotypes have a high probability of being different species.

3) Shells that look similar may belong to genera of diverging clades. Shell data provide the least valuable data for taxonomic discrimination above the species level, and especially above the generic level.

4) When the anatomical database is weak, one must rely more on molecular databases.

5) Genetic distances must be calibrated within and between clades in order to understand the magnitude of distance values that indicate specific status. Gaps between species groups and the direction of taxon divergence are clarified using multidimensional scaling and ordination with a network.

II: Population Genetics

Hybridization

The following example demonstrates the use of molecular genetics to uncover hybridization. Bianchi *et al.* (1994) electrophoretically examined the allozymes of two North American freshwater gastropod pleurocerid species, *Elimia virginica*, an eastern slope species, and *E.*

livescens, an interior basin species. These two species are readily distinguished by differences in shell sculpture. What drew our attention was the presence of both species in the Erie Canal, which connects the interior basin Lake Erie to the Hudson River, an eastern slope drainage system. The canal was built in 1825. Near the center of the canal at Mudlock, Strayer noted a blurring of shell phenotypes and asked the question: Could hybridization be occurring?

The electrophoretic study yielded data from 22 loci involving 39 alleles. Nei's genetic distance in the two species comparison was 0.332; Arc distance was 0.577, values indicative of species differences for North American Pleuroceridae (again,calibration) as evidenced by the work of Dillon (1984,1988), Dillon and Davis (1980), and Chambers (1978, 1980). We divided snails from the Mudlock location into two groups: one most resembling the *E. livescens* shell phenotype, with some shells appearing to be pure *E. livescens*, the other most resembling the *E. virginica* type with some shells appearing to be pure *E. virginica*. There were 11 polymorphic loci, of which nine were diagnostic. A sampling of results is given in Table 2. It is clear that hybridization and introgression have occurred. There are no F_1 generation snails involved as there is fixation for alternative alleles at the LAP locus. However, evidence for introgression was found at seven (78%) of the diagnostic loci.

In our study and those referenced above, mean heterozygosities are very low for populations and species (0.002–0.03 most common). Thus, the situation along the Erie Canal stands out prominently with H values of 0.035, 0.044. As is characteristic of hybrid zones, polymorphism increases, there is an increase in rare alleles, and heterozygote deficiency is frequently found (Hewitt, 1988; also note the presence of new electromorphs called hybrizymes by Woodruff, 1989). Vainola and Huilsom (1991) effectively described the occurrence of a hybrid zone for populations of *Mytilus*, but their diagnostic loci were not as distinct as those in our study.

Table 2. Sampling of allele frequencies from the pleurocerid hybridization study.

Locus	*virginica* control	*livescens* hybrid-mix	*virginica* hybrid-mix	*livescens* control
AAT-1				
100	1.00	0.097	0.711	—
97	—	0.903	0.289	1.00
GPI				
100	1.00	0.016	0.974	—
	—	0.984	0.025	1.00
NADD				
100	1.00	0.210	0.763	—
105	—	0.790	0.237	1.00
LAP				
100	1.00	—	1.00	—
97	—	1.00	—	1.00

Special Selective Pressures

In a study of populations of marine marshland snails, *Hydrobia truncata*, from North America, New England to Maryland, there were no significant anatomical differences except for size (Davis *et al.* 1988; 1989). Electrophoretic studies involved 30 loci, 49 alleles. There were 18 invariant loci. One population in a salt pond (Flax Pond) on Long Island, New York was of particular interest. This pond was 185 years old. The population differed from the others by Nei's D of 0.12 ± 0.01 while the other populations differed among themselves by a D of 0.07 ± 0.01. This difference is due, in part, to eight unique alleles. The F_{ST} statistic shows population differentiation due to the unique genetics of the Flax Pond population; there are no significant regional difference($F_{ST} = 0.004$). The Flax Pond population is different from the other populations for two additional reasons: 1) the snails are of gigantic size; 2) the snails are exceptionally heavily parasitized. The extreme parasite burden causes the gigantism (Davis *et al.*, 1988). The density of snails is also very high, some 25,000/m^2. The most plausible explanation for the maintenance of unique alleles in a population of only 185 generations is density dependent selection where any mutation yielding a unique allele yields some benefit to a population extraordinarily stressed by parasitism (an idea attributable to Haldane).

Breeding Structure

Allozyme electrophoretic studies play a central role in the modern study of genetic structure of populations and especially breeding structure. The topic has been reviewed as it relates to mollusks (Selander & Ochman, 1983). An example of such a study is provided by McCracken and Selander (1980), in which they studied the breeding systems of 14 species of three families of terrestrial slugs in the eastern United States. They demonstrated that for six species, the normal breeding system was either facultative or obligatory self-fertilization. One of the species had three monogenic strains. One of the most studied self-fertilizing species is the land snail *Ruminia decollata* with more than 30 monogenic or very weakly polymorphic strains (Selander & Kaufman, 1973).

Take Away Message

Allozyme electrophoresis is, and will continue to be, an essential tool for studying population structure, patterns of reproduction, and uncovering different effects of selective pressures.

III: Patterns of Evolution

A: Speciation

Species Definitions: One cannot speak of discriminating among species and engage in describing new species without committing to a species concept and being pre-

pared to defend the concept. This commitment is especially important if one is to apply molecular data in justifying species status. This is not the place, nor is there enough time in this presentation to argue at length all pertinent aspects for defending a particular concept. I will be brief.

There are essentially five concepts worth arguing about of which I accept the fourth in the list. These are extensively reviewed by Templeton (1989). (1) The Biological Species Concept defended by Mayr (1963) has also been called the isolation species concept (Patterson, 1985; Templeton, 1989). It is not acceptable as it stands because the emphasis is on isolating mechanisms. As most of us know from experience, the problem is not one of distinguishing species that occur in sympatry, but those that are allopatric, e.g. in Korea *vs.* Japan. Intrinsic isolating mechanisms are irrelevant as isolating barriers during allopatric speciation. This is not to reject the relevance of isolating mechanisms when they occur and are part of the process of speciation, such as an instantaneous cytological event within a population that causes reproductive isolation. However, the process of speciation in allopatry has nothing to do with "isolating barriers".

Because Mayr associated polytypic species with convenience in pigeon-holing taxa, various authors have rejected both the biological species concept and the polytypic species concept (e.g. Cracraft, 1989). However, these are two separate issues. There is value in understanding isolating mechanisms where they do apply during the process of speciation; and I agree with Templeton (1989) that polytypic taxa are relevant as a concept in theory and practice. It is a straw-man issue to dismiss both concepts just because some thought to use polytypic taxa as an excuse to reduce the number of names one would have to deal with. The polytypic species issue will be revisited later.

Gareth Nelson (1989) rejects the biological species concept because it assumes that species are the basic units of evolution. Since he rejects the idea that species are the basic units of evolution, he rejects the concept of species! While I can agree that species are not the basic units of evolution, I still can accept the reality of species. Populations are the fundamental units of evolution in as much that mutations or new gene combinations in individuals of a population are basic to the process of speciation, and excepting clonal and/or selfing individuals, population structure is necessary for the spread of new genes.

I reject both the evolutionary (model 2) and phylogenetic (model 3) species concepts for the same reasons. Wiley's (1981) evolutionary species concept involves "a lineage which maintains its integrity from other such lineages and has its own evolutionary tendencies and historical fate". It requires parental system of ancestry and descent. It requires large disjunctions in allopatry where one or more unique diagnostic qualitative characters are needed. The phylogenetic species concept of Rosen (1978) and Cracraft (1989) is the same but the analysis must be done by cladistic methods. "A phylogenetic species is an irreducible (basal) cluster of organisms, diagnosably distinct from other such clusters, and within which there is a parental pattern of ancestry and descent". As Nelson and Platnick (1981) stated earlier, "species are simply the smallest detected samples of self perpetuating organisms that have unique sets of characters". At this point one should note that *all* species concepts discussed here involve taxa that are monophyletic, i.e. there is a parental pattern of ancestry and descent; even the biological species concept.

The problems with the above concepts (models 2 and 3) are that they provide no guidance as to what characters are important for defining species. By their criteria, two allopatric populations that differ in an electrophoretic study of allozymes, by two or three alleles unique to each population, should be called different species. Hardly! While both concepts admit to cohesion, i.e. the integrity of a morphological database, they do not allow for variance, or how much variance can be associated with a species. They deal only with the concept of cohesion, not with the mechanisms responsible for cohesion. Ignorance of population genetics as it relates to expression of multiple alleles at a single locus would easily lead to splitting populations into discrete species when they really should be considered one species.

I prefer the cohesion species of Templeton (1989) (model 4) that includes the recognition concept (model 5) of Patterson (1985). According to Patterson, a species is a population of biparental organisms sharing a common fertilization system. This is the "flip side" of the biological species concept of Mayr.

For Templeton, speciation is the evolution of cohesive mechanisms. A species is a group of populations of a monophyletic lineage. There is phenotypic-reproductive cohesion. **These populations share the same fundamental niche, i.e. the populations are interchangeable = demographic exchangeability**. Natural selection promotes cohesion both through favoring reproductive cohesion, genetic relatedness and affecting the limits of demographic exchangeability. The concept integrates population genetics and ecology with the standard studies of morphology. The concept can be applied to all organisms from outbreeders to syngameons or parthogenetic organisms.

Using this concept requires more work and more data, but who ever said that understanding the process of speciation was simple and uncomplicated? The process of distinguishing species is not simply akin to picking up marbles and assigning names to them on the basis of size, surface patterns, and colors. With the cohesion concept, two populations, one in Japan and the other in Korea, might have their own evolutionary fates ahead of them, yet still belong to the same species.

Patterns of Speciation: The relative ease of discriminating among species very much depends on the pattern of speciation one encounters. There are two major modes that I have encountered again and again: adaptive radiation and morphostatic radiation. Osborn (1918) first

used the term adaptive radiation: " Adaptive radiation is, descriptively, this extreme diversification of a group [e.g. mammalian, reptilian] as it evolves in all the different directions permitted by its own potentialities and the environments it encounters." For Stanley (1979), adaptive radiation is the rapid progression of new taxa from a single ancestral group. If one takes Stanley's definition, one might be dealing with the concept that Osborn had in mind, where there is a considerable diversification in morphological ground plans, or one might be involved with a morphostatic radiation as defined by Davis (1992).

Benton (1988) very well captured the essence of the Osborn concept. The "key phases of an adaptive radiation such as that of the placental mammals 65 million years ago . . ." involves (1), "an initial phase of rapid diversification from a single ancestor. . . (2) the establishment of a diversity of new body plans in this early phase. . . (3) early extinctions amongst the initial elements of the radiation. . .(4) a final phase of stabilization of the lineages. . . ." It is important to assess the time scale of an adaptive radiation. The mammalian radiation exploded between 65 to 55 million years ago. All the orders from bats to whales were established in this 10 million year period. There is also a taxon scale to consider. There has been extraordinary morphological diversification within the freshwater gastropod family Pomatiopsidae, but especially in the subfamily Triculinae that is wholly southeast Asian and southern Chinese. In relatively short time, some 12 million years to the present, these small snails have diversified into three tribes and over 23 genera (Davis,1992). The rapid diversification probably took place in the first few million years (Davis *et al.*, 1984). There is a splendid adaptive radiation that centers in the Mekong River (Davis, 1979) that involves some ten of the 23+ genera. There is also a large morphostatic radiation in southern China.

I use the term morphostatic radiation to include those monophyletic taxa that have indeed radiated, but in allopatry and where there are little or no discernible niche differences. Likewise there is comparatively little morphological differentiation compared to what one sees in an adaptive radiation where morphology may vary in a number of dimensions that reflect adaptations to different environments. I use the term to replace Gittenberger's (1991) "non-adaptive radiation" that is: ". . .non-adaptive radiation should denote evolutionary diversification from a single clade, not accompanied by relevant niche diversification. . .the various species resulting from the process would,in principle, not be able to be sympatric." As it is reasonable to consider that all species in nature are adapted to their environments, the term non-adaptive seems unsuitable. Further, one can discern between adaptive radiations *sensu* Osborn, where morphological adaptations to differing environments are considerable, and morphostatic radiations as defined above. Gittenberger maintains that there are numerous intermediate situations between the two types I describe, and that the concepts of higher taxa and genera are too ill defined. I reject these notions. There are sufficient quantitative and cladistic procedures to clearly define taxa except in certain instances of morphostatic radiation where the mosaic of few characters confounds understanding the limits of species in allopatry.

Table 3. Scoring genera and the radiation of species of each genus for qualitative morphological and ecological differences involving eight characters that includes ecology. See text for details. A_d = coefficient of radiation diversity.

Adaptive radiation: Triculinae	*Hubendickia*	*Lacunopsis*	*Pachydrobia*	*Jullienia*
Shell	5	10	10	10
Radula	5	5	0	0
Mantle cavity	0	5	0	0
Head	0	5	0	0
Female reprod. system	5	5	5	5
Male reprod. system	5	0	0	5
Nervous system	0	0	0	0
Ecology	5	10	10	5
Sum scores =	25	40	25	25
A_d = 3.68				

Morphostatic radiation: Triculinae	*Tricula*	*Neotricula*	*Gammatricula*	*Wuconchona*
Shell	0	0	0	5
Radula	0	0	0	0
Mantle cavity	0	0	0	0
Head	0	0	0	0
Female reprod. system	5	5	5	5
Male reprod. system	0	0	0	0
Nervous system	0	0	0	0
Ecology	0	0	0	0
Sum scores =	5	5	5	10
A_d = 0.82				

In a morphostatic radiation one may be able to discern among species or subspecies because of shell sculptural differences such as seen in *Albinaria* of Gittenberger's (1991) example. There may be a mosaic of quantitative differences that separate species. But these are, in such radiations, usually small differences. To make the point, I compare four genera of the Triculinae adaptive radiation with four genera of the Triculinae morphostatic radiation in Table 3. Each genus is scored for qualitative differences in morphology and ecology compared with other genera, where the species occupy different ecologies. There are eight characteristics scored. A genus is scored 5 when it clearly differs from other genera in a change in ground plan; species in the genus are scored 5 when they diverge in a character. For example, *Lacunopsis* differs from other genera in ecology (5) and the species have radiated into different niches (5); ecology thus scores 10. In the same genus the radula differs from others in the tribe but the radula is the same in all of the species, therefore it scores 5 for radula. In a morphostatic radiation one cannot tell the shells of *Neotricula*

from those of *Tricula*; shells score 0 as the species do not radiate with different shell shapes or sculpture; etc.

To quantify the differences between the types of radiation, I use a coefficient of adaptive differentiation, A_d, that is calculated using the formula:

$$A_d = \sqrt{\frac{\left(\frac{\Sigma S_1}{n_1}\right)^2 + \left(\frac{\Sigma S_2}{n_2}\right)^2 \cdots \left(\frac{\Sigma s_x}{n_x}\right)^2}{N}}$$

It is a standardized sum of mean squared differences where n = number of characters scored; $S_1S_2 \ldots S_x$ =individual score for a character for generic species group 1, etc.; N = number of generic species groups. From the example given in Table 3, there is 4.5 times the amount of diversification morphologically and ecologically in the adaptive radiation of triculine taxa compared to the morphostatic radiation of triculines.

I have digressed considerably on species definitions and patterns of speciation because it becomes clear that discriminating among species in an adaptive radiation, and where several congeneric species may be located in a habitat, may be a relatively easy task. The difficulties reside with allopatric taxa of a morphostatic radiation where molecular genetics may be extremely useful or may confound the issue. The situation I presented above for the three species of *Uniomerus* is an example of usefulness of molecular genetics. I will present two examples where the interpretation of molecular genetics must be made in light of all other data.

***Oncomelania hupensis* polytypic species:** *Oncomelania* is a member of the Pomatiopsidae: Pomatiopsinae, with what I currently consider to be two species: *Oncomelania hupensis* and *O. minima*. This genus is a member of a morphostatic radiation. The latter species occurs in Japan; the former has subspecies distributed in China (1), Taiwan (2), Japan (1), Sulawesi (1), and the Philippines (1). The systematics of this genus was reviewed by Davis (1980,1981). Briefly, *O. hupensis* was considered to be a polytypic species because: (1) the anatomy, except for differences in size, was identical for all allopatric populations; (2) the main shell differences were the occurrence of ribs on some populations in southern China; however ribbing is controlled by a single gene and this gene has been manipulated by hybridization experiments; (3) the populations can be hybridized with no loss of viability of F_1 or subsequent generations; (4) as anyone who has simultaneously raised these snails in culture will attest to, there is ecological exchangeability between populations. There are indeed small differences in susceptibility to different allopatric populations of *Schistosoma japonicum*; in size; in degree of shell varix formation; and the degree of gland formation about the medial aspects of the eyes.

Recently, Woodruff *et al.* (1988) did an electrophoretic allozymic study of several populations of *Oncomelania hupensis quadrasi* from the Philippines and compared them with *O. h. hupensis* from China. The Nei's D among the Philippine populations averaged 0.036 (greatest value = 0.134); The Chinese and Philippine populations differed by a mean of 0.62 ± 0.04. They concluded that the Chinese and Philippine snails belonged to different species because of the large genetic distance. They justified this conclusion by (1) invoking the evolutionary species concept; (2) they reject the biological species concept with its emphasis on reproductive isolation and consider polytypic species to be an essential element of the biological species concept; (3) they argue that subspecies is a category of convenience, a way of pigeon-holing taxa and reducing the number of names one has to deal with; (4) they argue that the use of subspecies causes confusion by underestimating the number of independently evolving lineages.

I disagree! The polytypic species concept with its subspecies is indeed useful; it certainly does not have to be married to the biological species concept. Also, the recognition of subspecies does not reduce the number of names used nor does it imply that the allopatric populations involved are not independent evolving lineages. The use of subspecies serves a very useful purpose and is not used as a matter of convenience. Further, the *Oncomelania hupensis* polytypic species complex does meet the major criteria of the evolutionary species concept: reproductive recognition and genetic integrity.

Large genetic distances by themselves do not serve to define species. For example, Johnson *et al.* (1984) found that a population of *Cepaea nemoralis* introduced from Europe to Lexington,Virginia (southern U.S.A.) differed from a population from Florence, Italy by Nei's D of 0.631; a population from Santa Croce near Pavia, Italy compared with the Florence population differed by D of 0.391. This species is well known for geographic variation for both shell polymorphism and allozymes. The species is well studied throughout its range. To quote Johnson *et al.* (1984), "The decoupling of genetic divergence from speciation emphasizes the limitations of viewing the process of speciation solely in genetic terms." In questioning the origin of the Lexington population, Stine (1989) used restriction enzyme analysis of mitochondrial DNA to demonstrate it to be more closely related to populations from England rather than from Italy (Nei's D of 0.409).

What Woodruff *et al.* (1988) are doing is ignoring the great genetic cohesion that unites the subspecies of *Oncomelania hupensis*, a cohesion that is associated with demographic exchangeability. One powerful example of this genetic cohesiveness is the invariability of the reproductive systems. There is indeed cohesion in reproductive recognition. While I certainly agree that the ability to hybridize is not a criterion for merging perfectly good species (examples of syngameons are numerous), it is instructive that the large Nei's D does not interfere with the hybridization of these subspecies with no loss of viability in the offspring or through successive generations. There is indeed genetic cohesiveness. Be-

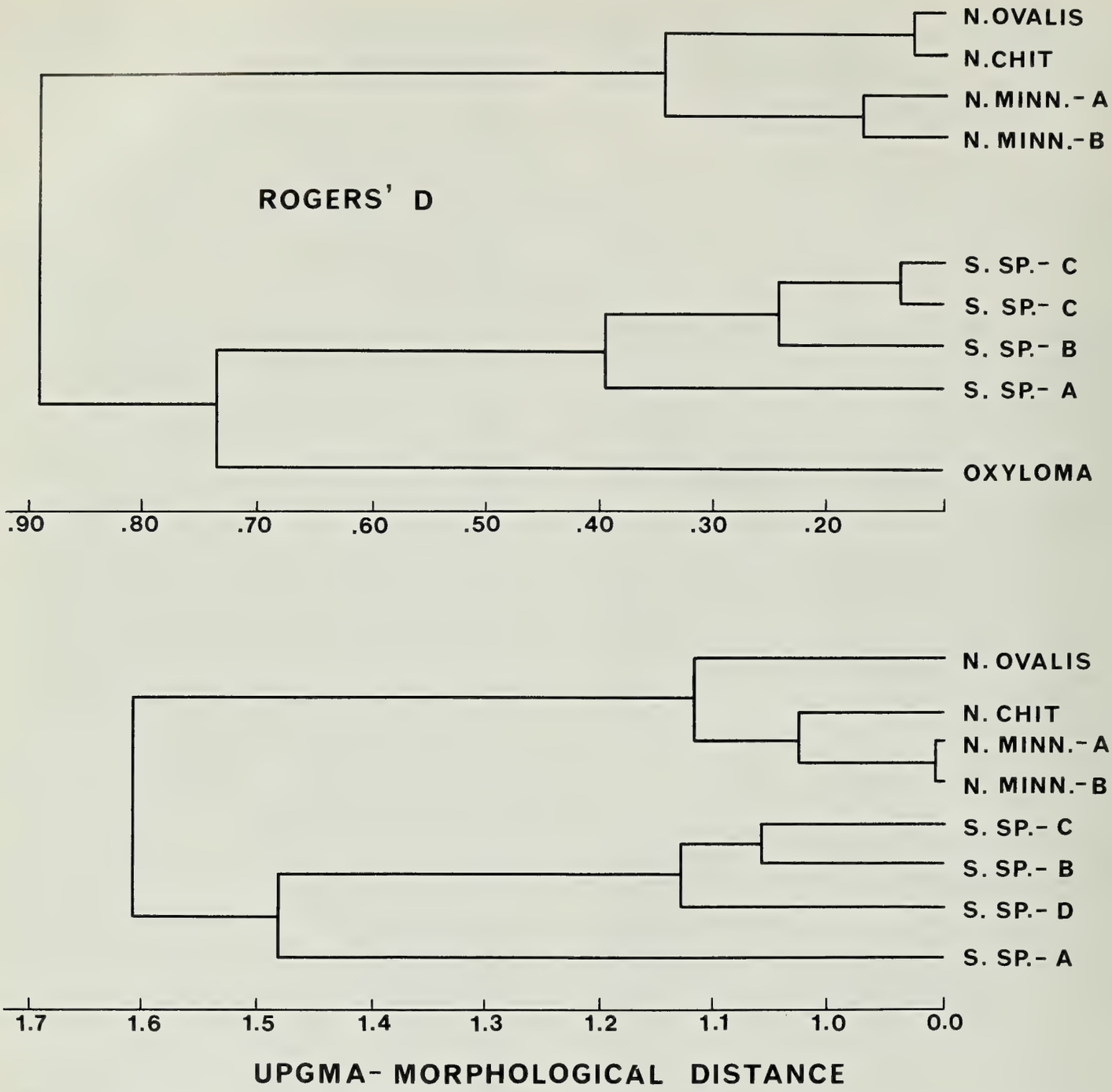


Figure 5. A comparison of phenograms comparing some of the succineid taxa studied by Hoagland and Davis (1987). Allozymic electrophoretic results are compared with morphological results. Modified from Hoagland and Davis (1987). Sp. D was not studied electrophoretically. S= *Succinea*; N=*Novisuccinea*. See text for details.

yond the cohesiveness of the reproductive organs, there is the cohesiveness in all the other details of anatomy, reproductive habit, responses to environmental manipulation. The usefulness of the subspecies designation in this example is to bring attention to the great cohesion throughout this complex and understand what this implies for many aspects of the biology of the species through time. It would be useful if Woodruff *et al.* would examine the disruption of cohesion (i.e. morphological diversification) in sister taxa to *Oncomelania* that are likewise considered part of a morphostatic radiation, i.e. species of *Tricula* or *Neotricula*. One finds numerous characters of use to distinguish among species. Examples of these are most frequently found in slight modifications of the reproductive systems; e.g. penis with papilla in one species, without papilla in another; penis with pronounced ejaculatory duct in one species, without ejaculatory duct in another; penis mounted center on the head *vs.* right of center; seminal receptacle arising at position "a" in one species, or in position "b" in another. Also,the shell

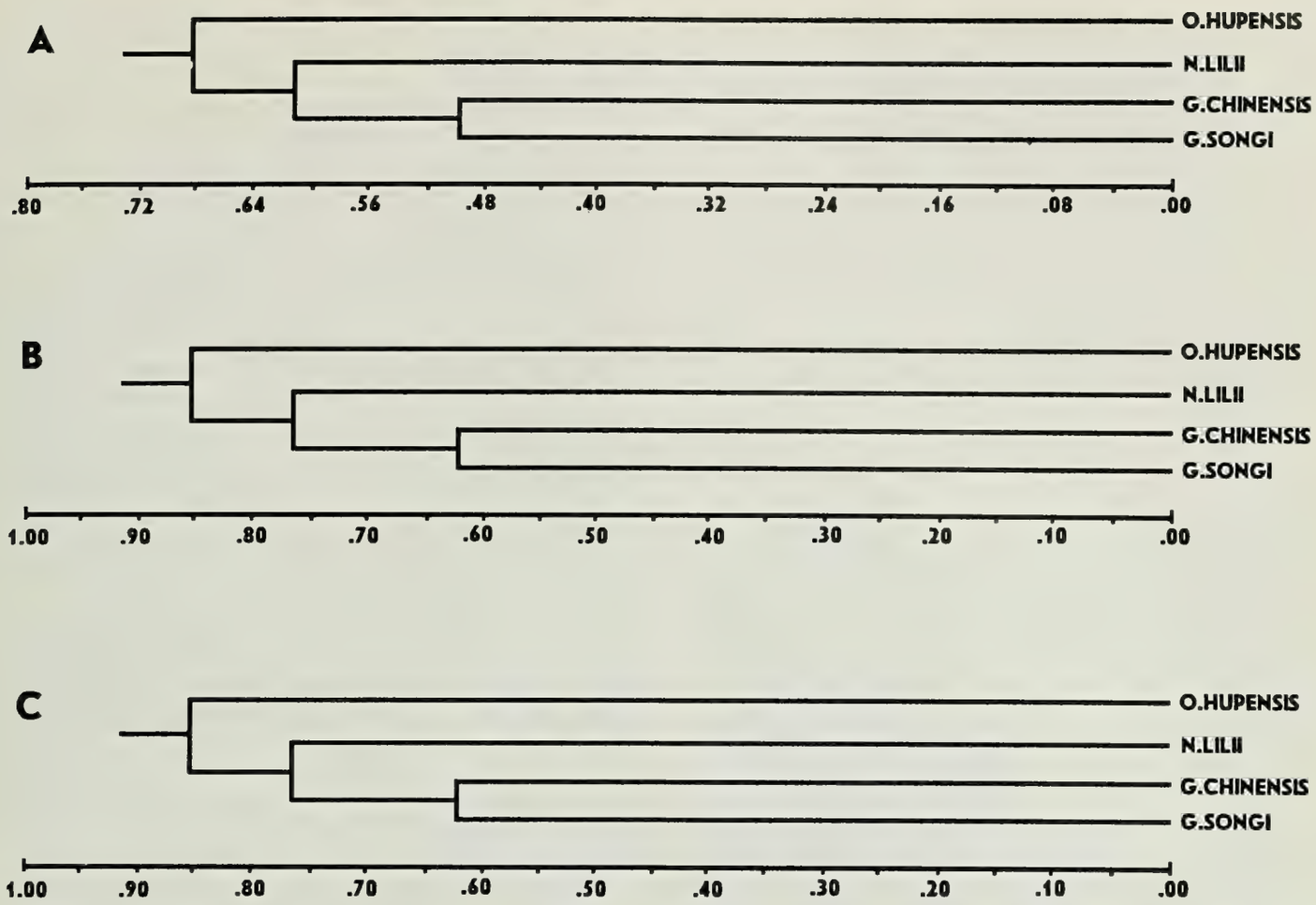


Figure 6. Phenogram based on UPGMA treatment of genetic distances. **A.** Nei's D; **B.** Wright's modified Rogers' D; **C.** Arc D. O. = *Oncomelania*; N.= *Neotricula*; G.= *Gammatricula*. Adapted from Davis *et al.* (1994b).

may have an internal tooth on the columella *vs.* no tooth; and so on. These small differences are the types found in a morphostatic radiation in contrast to major ground-plan changes found in an adaptive radiation. The contrast with polytypic *Oncomelania hupensis* is striking. To achieve full species status, some occurrence must cause disruption of the cohesion seen. This is evidenced in *Oncomelania minima* of Japan where the shell shape departs from that seen in *Oncomelania hupensis*; there

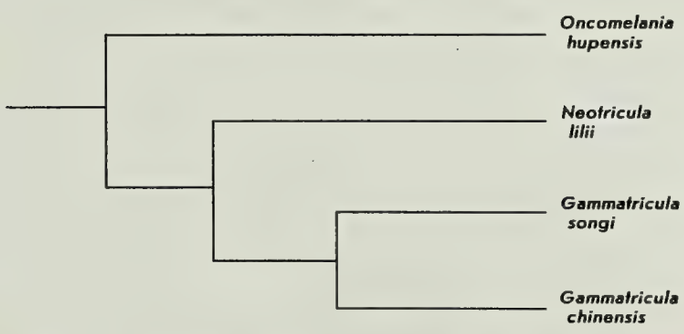


Figure 7. Cladogram based on a Hennig86 treatment of the same allozyme data used in Figure 6, but scoring each locus as a character.

are several shifts in the morphologies of the reproductive systems (Davis, 1969), character-state changes similar to those seen in the sister subfamily Triculinae. The point is that one must know what occurs in sister taxa relative to the taxon under study; what are the patterns of character change relative to recognition of species.

Unfortunately, the use of subspecies in malacology is generally farcical! No wonder the term subspecies is little respected when subspecific status is awarded to populations that differ by so slight a character-state as an extra bump or node or rib in one population that is not seen in another. Numerous subspecies have been based on conchology alone where the basic definition of a species has not been worked out, let alone any understanding of what the extra bump means. However, while most malacological subspecies currently named in the literature have no biological validity, there are indeed substantiated cases of polytypic species, and *Oncomelania hupensis* is one of them.

The land snail genus *Succinea*: In this example, comparative anatomy, ecology, and molecular genetics were used to assess species status. As will be shown, molecular genetic data were useful in some cases, not useful in other cases. The question was, what was the true identity

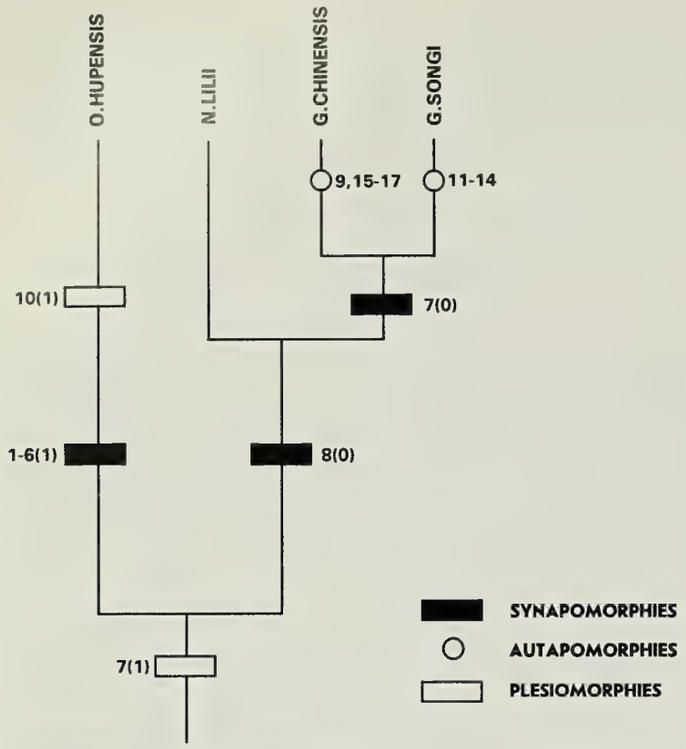


Figure 8. Cladogram based on morphological data for the same taxa shown in Figures 6, 7.

of a rare and endangered species, *Succinea chittenangoensis* Pilsbry, and how could it be distinguished from other sympatric species on or at the Chittenango Falls in upper New York State in the northeastern USA? The study included topotypical *S. ovalis* and populations of *Succinea* from Pennsylvania and Minnesota. The outgroups for the electrophoretic study were *Oxyloma retusa* Lea, and *O. decampi gouldi* Pilsbry of Chittenango Falls. Through electrophoresis and shell morphometrics, another species was found at the falls in addition to *S. ovalis* and *S. chittenangoensis*, a species with a shell shape similar to *S. putris* (Linnaeus) of Europe. In the anatomical studies, 51 characters were scored using binary coding. In the allozyme studies, 31 loci involving 87 alleles were found. The data of both sets were analyzed using multivariate analysis yielding UPGMA derived phenograms as shown in Figure 5 [modified and simplified from Hoagland & Davis (1987)]. The findings were: (1) Genetic and morphological data support the conclusion that three genera are involved; *Succinea*, *Novisuccinea*, and *Oxyloma*. (2) *Oxyloma* is more closely related to *Succinea* than it is to *Novisuccinea*. (3) *N. ovalis* and *N. chittenangoensis* at the falls cannot be distinguished electrophoretically while they are clearly distinct in terms of anatomy and ecology (as well as on shell differences). They are distinct species. (4) The two populations of *Novisuccinea* from Minnesota are clearly not *N. ovalis*. They are not morphologically distinct yet they have diverged genetically (Nei's D = 0.104). Further studies would be necessary to assess whether or not they are specifically distinct. (5) In the remainder of the comparisons results based on morphology paralleled those based on molecular genetics.

Why are the falls *Novisuccinea* species morphologically divergent yet not electrophoretically so? The probable answer is that the area was glaciated until 10 to 12 thousand years ago. With retreat of the glaciers and the uncovering of the falls, *N. chittenangoensis* evolved from an ancestor of regional *N. ovalis* by colonizing the falls with concomitant shifts in morphology in adapting to new ecological space. There has not been enough time to diverge in terms of allozymes.

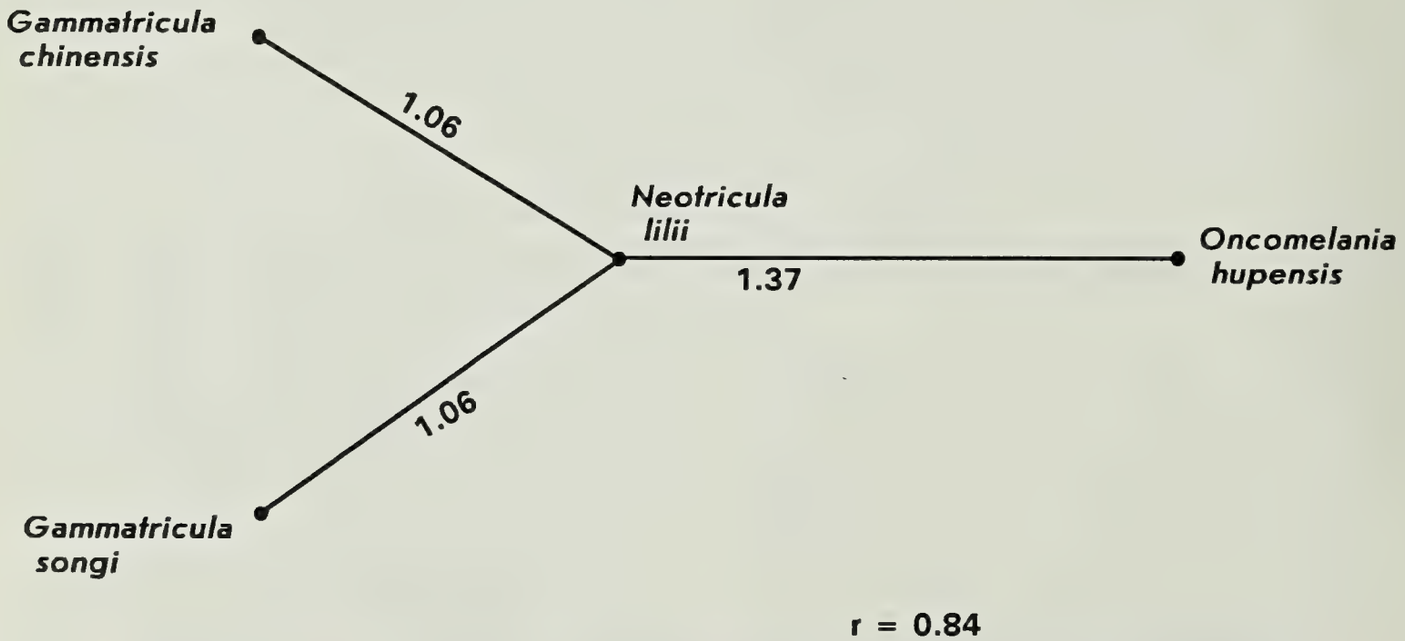


Figure 9. Prim network based on multivariate analysis of morphological data. The taxa are those treated in Figures 6–8.

B: Phylogeny

Molecular genetics are certainly useful in assessing phylogeny. I will provide an example that builds on the morphostatic radiations given above involving the Pomatiopsidae: Pomatiopsinae and Triculinae (Davis *et al.*, 1994b). The questions asked were: Are the Pomatiopsinae and Triculinae monophyletic? Is *Oncomelania* closely related genetically to the more generalized triculine taxa that are part of the morphostatic radiation? What genetic distances might one expect between genera of the Triculinae? Is a cladogram based on genetic data congruent with a cladogram based on anatomy? Are these cladograms congruent with biogeographical data?

The electrophoretic analysis involved 28 loci and 78 alleles. The morphological analysis involved 17 characters. Phenograms based on UPGMA treatment of three genetic distances are given in Figure 6. A cladogram based on using each enzyme locus as a character and applying Hennig86 version 1.5 (Farris, 1989) is given in Figure 7. *Oncomelania* is the outgroup; there was no differential weighting or polarities assigned. Only one tree resulted, with a consistency index of 0.62. The cladogram based on the morphological data is given in Figure 8. The phenograms and the two cladograms are congruent. The cladograms are congruent with biogeography and the hypothesis on the direction of evolution from northern Burma-western Yunnan, China with dispersion and divergence down evolving river systems (Davis, 1980, 1992). These congruencies give confidence about the phylogenetic results published earlier based solely on comparative anatomy (e.g. Davis & Kang, 1990; Davis, 1992)

The question about monophyly is also answered. There is no great divergence of *Oncomelania* from the triculine taxa. The question was justified for the following reason. In the Pomatiopsinae, the spermathecal duct runs from the bursa copulatrix to the anterior end of the mantle cavity. In the Triculinae the spermathecal duct runs from the bursa to the pericardium or to the posterior end of the mantle cavity. Are the spermathecal ducts homologous? It has been a hypothesis that the spermathecal duct in the Triculinae derived from the primitive gonopericardial duct that connects the oviduct to the pericardium in some rissoacean taxa. However, in the Pomatiopsinae there is a vestigial gonopericardial duct and the spermathecal duct! Two families might be involved.

The average Nei's D between *Oncomelania* and the triculine taxa is 1.29 ± 0.41. This is not a large distance considering what one might expect of different families. It is especially not large when one calibrates the system. The distance between two of the triculine taxa is 1.26, a D value greater than between *Oncomelania* and *Gammatricula songi* where D = 1.00. In describing *G. songi*, Davis *et al.*(1994) stated that the anatomical innovations found in this species warranted generic status, but that a new genus would not be named until more species of *Gammatricula* were found and studied. As shown in Figure 9, *G. songi* and *G. chinensis* diverge equally from *Neotricula* along the Prim Network. Considering there to be three triculine genera involved, the average Nei's D among them is 0.890 with a range of 0.689 to 1.236. Thus, *Oncomelania* seems more to be a genus closely allied within a triculine generic grouping rather than a member of a different subfamily. Once again the point is made: Measures of genetic distance do not serve to define taxon levels! The subfamilies Pomatiopsinae and Triculinae are firmly based on qualitative anatomical data that in either a phylogenetic/ cladistic or multivariate analysis support those diverging sets of genera at a hierarchical level deserving subfamilial status. The genetic data do serve to confirm close genetic relationship, not a highly disjunct pattern indicating polyphyly.

Take Away Message

Discriminating taxa at the species level is most difficult when one is dealing with allopatric populations of a morphostatic radiation. There are indeed different processes of speciation. Speciation may proceed uncoupled from genetic differentiation seen in structural genes such as demonstrated using allozymes. Considerable genetic distances do not necessarily mean that the overall genetic cohesiveness among populations is disrupted to the extent that species status is attained. Rapid morphological change in adapting to new environmental space may outpace molecular genetic change. In examining a large radiation spread over great distances, one would expect that in perhaps 70% or more of the species, morphological and molecular genetic change would diverge in parallel. Untangling species-level problems can be a most challenging task as pointed out by Giusti and Manganelli (1992; see Prologue). For those engaged in this task, one needs as much data as one can obtain, certainly building on a firm platform of detailed comparative anatomy. Ecological data are essential. Molecular data are always useful, but do not add to the solution of a problem in a rote formulated way. Above all, molecular data must be calibrated for the radiation under study.

Concerning phylogeny: Molecular genetic tools are essential to test phylogenies based on comparative anatomy. Together, both data sets provide insight into the rate of evolution. Together, both data sets serve to test hypotheses about biogeography.

IV: Uncovering Unique Aspects in Evolutionary Process

It has been known for a long time now that different molecular data sets may yield different results. Also, one set of tools is better suited for assessing relationships at one taxonomic level, while other tools are better suited for a different taxonomic level. For example, restriction enzyme analysis of mitochondrial DNA is most suited for determining relationships at the population level, or among closely related species. Allozymes are superb for studies of population genetics and to assess relationships

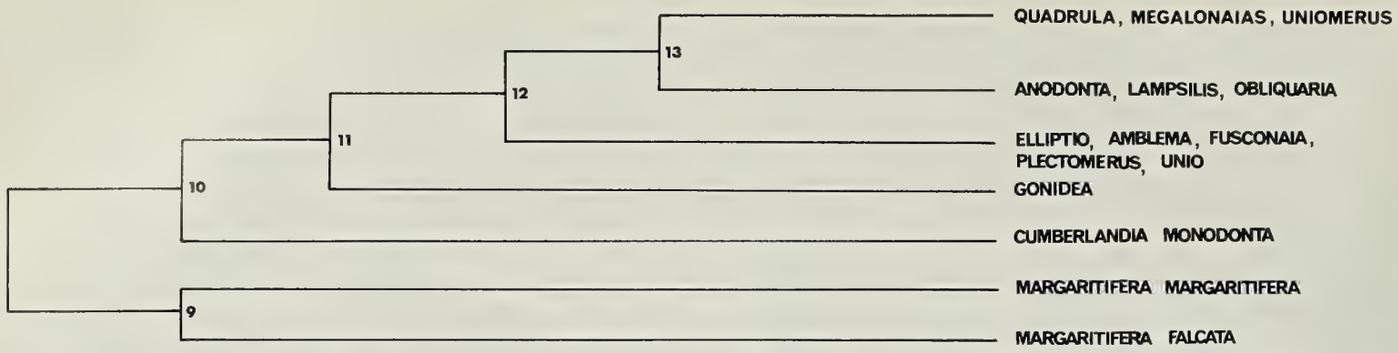


Figure 10. Phenogram following UPGMA treatment of distance coefficients based on LrRNA sequence differences among species of freshwater clams of the family Unionidae (for further details, see Rosenberg *et al.*, 1994).

among species, monophyletic genera and tribes to subfamilies. Immunology has been used with success from the species to family level. However, as discussed above, there is no universal molecular clock. Different data sets may yield different results. For example, Murray *et al.* (1991), as part of a series of excellent studies of evolution and speciation within the Pacific islands land snail *Partula*, compared the results of morphological investigations, protein electrophoresis, and mtDNA and defended the following: ". . .the different data sets evolve independently and at variable rates. This mosaic pattern of evolution can only occur if natural selection plays a role in the genetic differentiation of *Partula*".

We are now in a new age, one of sequencing. Cloning genes and nucleic acid sequence analysis began to explode in the decade of the 80's. Exciting developments were made possible with DNA amplification by the polymerase chain reaction (PCR) where a DNA segment of some 6000 base pairs may be amplified starting with as little as a single gene copy (reviewed by Landergren *et al.*, 1988). As with the emergence of any new technology, one should expect some surprises. Paradigms based on studies of mammals may be shattered when studying mollusks that have evolved over 600 million years with amazing diversification of anatomical groundplans, physiologies, and genetics. I present one such surprise encountered when studying large-ribosomal-RNA sequences of a series of freshwater clams (Unionidae) as part of a larger study that included land snails and the prosobranch *Oncomelania* (Emberton *et al.*, 1990).

The taxa studied are listed in Table 4 in the classification scheme of Davis and Fuller (1981). We examined some 150 base sequences that included the highly conserved 5′ end and the D–6 divergent domain plus flanking regions. We scored 26 differences among taxa and subjected these to a simple standard UPGMA treatment with the resulting phenogram shown in Figure 10. I wish to make only a few remarks about these results; a more detailed treatment of these sequence data in relationship to sequence data from diverse mollusks is presented later in this issue (Rosenberg *et al.*, 1994).

Morphological, immunological, and allozyme data support the concept that there are three equal and divergent clades (Figures 11, 12, adapted from Davis & Fuller,1981; Davis *et al.*, 1981). I would prefer to call them subfamilies, while others have split off the group of *Margaritifera* as a separate family. I point out, however, that given the weight of evidence, the group of *Anodonta* is equally divergent from other non-*Margaritifera* unionids and thus should be accorded equal rank either at the family level or subfamily level.

While there is congruence of the morphological,allozymic, and immunological data, the LrRNA se-

Table 4. Unionid species used to study LrRNA sequences classified in the scheme of Davis and Fuller (1981) based on immunological and morphological data.

Margaritiferinae
Cumberlandia monodonta
Margaritifera margaritifera
Margaritifera falcata
Anodontinae
Anodonta cataracta
Anodonta imbecilis
Anodonta grandis
Ambleminae
Gonideini
Gonidea angulata
Pleurobemini [should be Unionini]
Elliptio complanata
Pleurobema cordatum
Fusconaia cerina
Unio pictorum
Uniomerus "tetralasmus"
Amblemini
Amblema plicata
Quadrula quadrula
Quadrula cylindrica
Megalonaias boykiniana
Plectomerus dombeyianus
Lampsilini
Lampsilis claibornensis
Lampsilis teres
Obliquaria reflexa

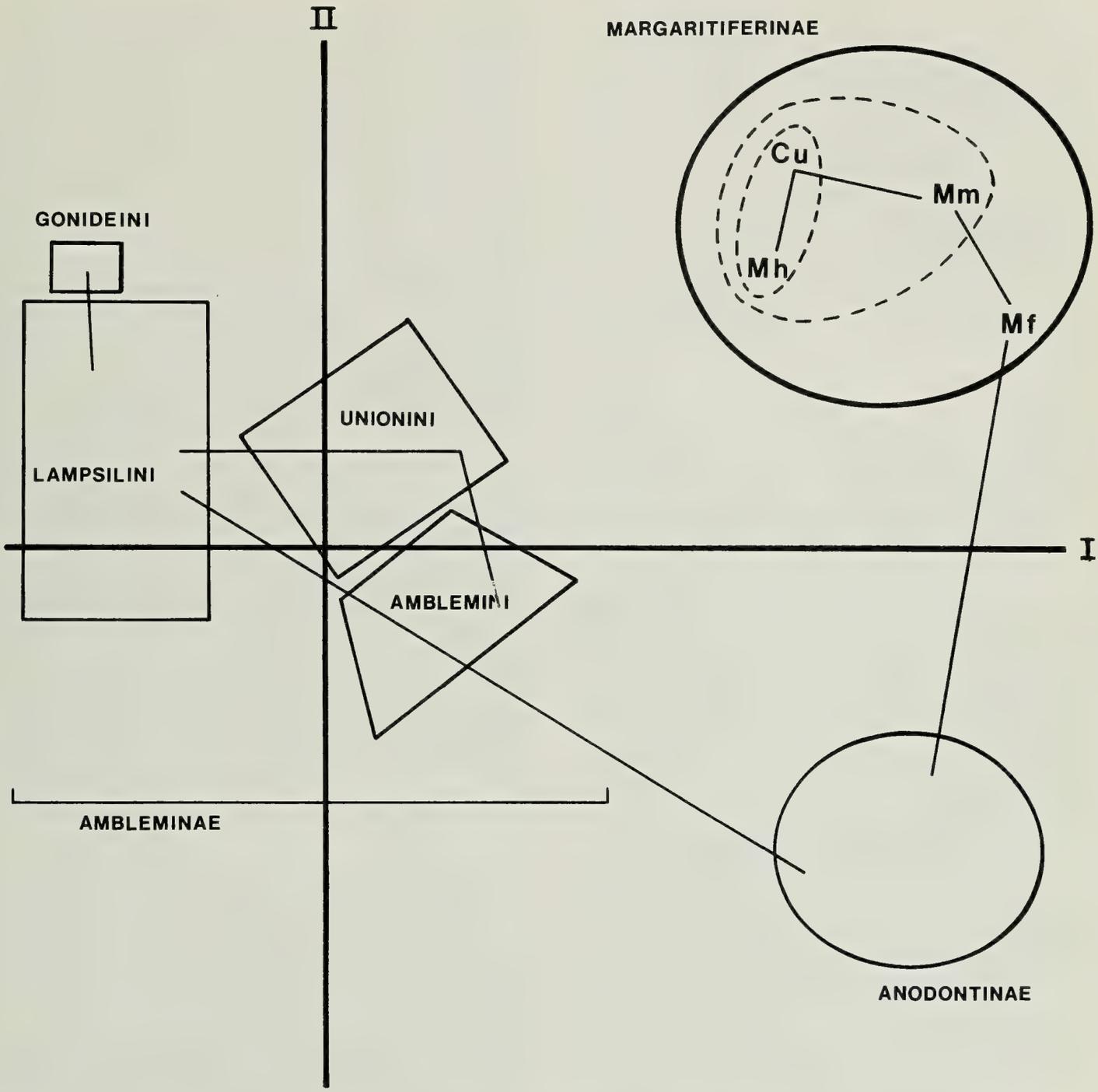


Figure 11. Ordination diagram following multidimensional scaling using immunological distances from freshwater clams of the family Unionidae. Computer-derived sets and subsets are enclosed in the dashed lines. As with the allozymic data-set (Fig.10), there are three discrete clusters: Margaritiferinae, Anodontinae, and Ambleminae(Amblemini,Lampsilini, Unionini, Gonideini). Adapted from Davis and Fuller (1981).

quence data offer some surprises! As seen in Figure 10, (1) *Anodonta* cannot be distinguished from Lampsilini genera *Lampsilis* and *Obliquaria*. (2) One cannot distinguish among species or genera in the groupings of *Anodonta* etc., *Elliptio* etc., and *Quadrula* etc., yet there are distinct differences between the two species of *Margaritifera*. (3) *Cumberlandia monodonta*, is widely separated from the species of *Margaritifera*. The differences among the Maragitiferinae taxa and with other unionids involve sequence changes at 25 positions, while the difference between the *Anodonta* group of genera and the *Quadrula* group of genera involves only one difference. The point to be made here is that *Anodonta* seems firmly nested with other genera of the tribes Unionini ,Amblemini, and Lampsilini while on all other data, both morphological and molecular (immunological and allozymic), *Anodonta* is highly divergent from genera of those tribes (Figures 11,12). The three Margaritiferinae taxa

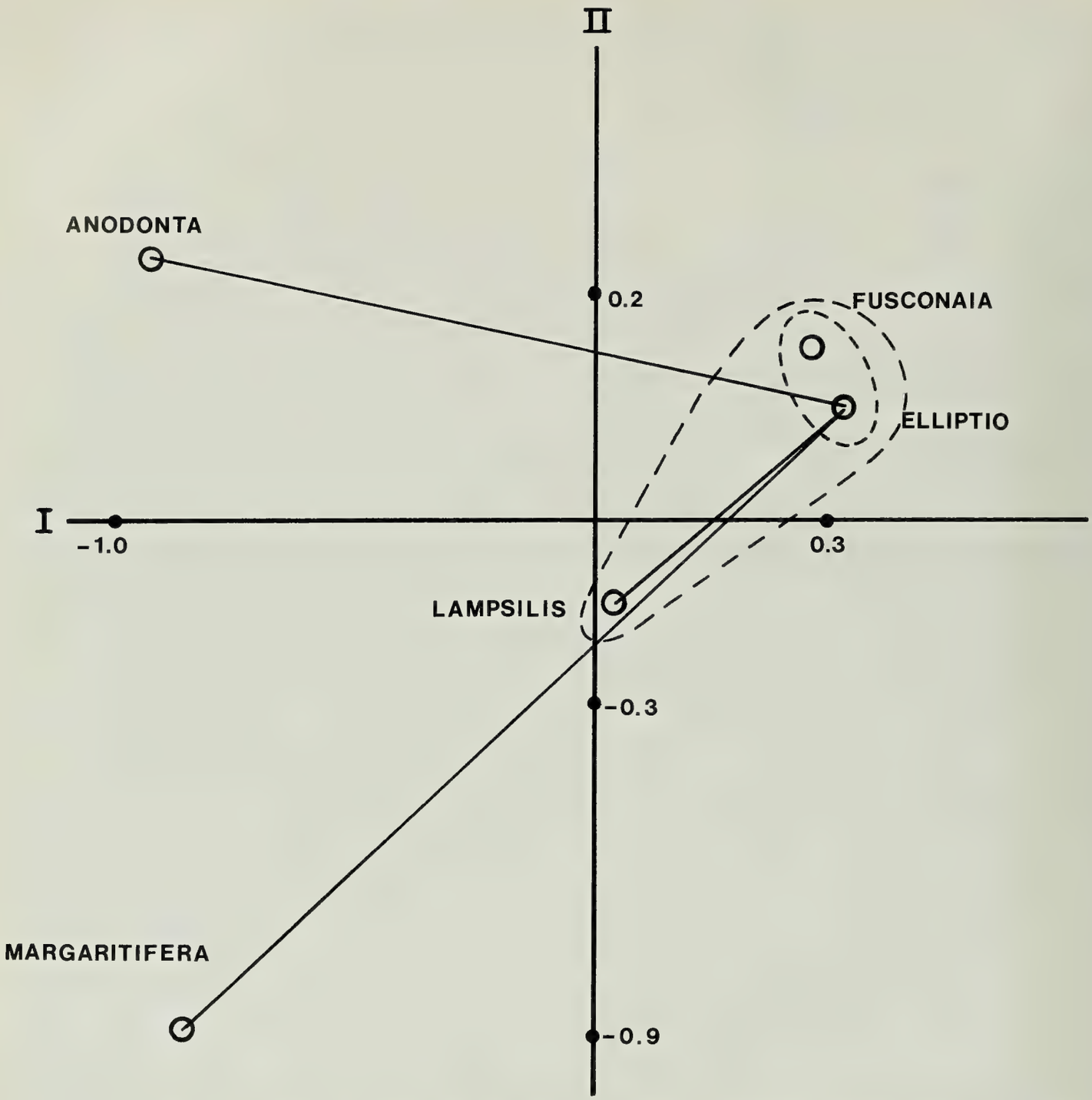


Figure 12. Ordination diagram following multidimensional scaling using allozymic electrophoretic data from freshwater clams of the family Unionidae. Computer-derived sets and subsets are enclosed in dashed lines. A Prim network is used to connect taxa. Note the direction of divergence of *Anodonta* away from *Margaritifera*, approximately equidistant from the set of the Ambleminae (*Lampsilis*, *Fusconaia*, *Elliptio*). Adapted from Davis *et al.* (1981).

show considerable divergence among themselves, with considerable changes in the variable and 3' flanking region, not seen in the other unionids. This is indeed a surprise. One interpretation that warrants further testing, is that the Margaritiferinae diverged from all other unionids at an early date and uniquely departed from other unionids in this pattern of sequence changes. *Anodonta*, while maintaining the rather conservative sequence structure, diverged from the non-Margaritiferinae clade, also at an early date, and rapidly diversified morphologically with concomitant immunological and allozymic changes.

Take Away Message

It is clear that a single measure of genetic distance cannot be used to discriminate among taxa. Speciation indeed

progresses by different patterns and processes. Speciation may proceed unhinged from genetics (as evidenced by current molecular techniques); some allopatric populations may retain great cohesion in breeding system, morphology and demographic exchangeability (all genetically controlled) yet accumulate considerable structural gene changes that, by themselves, do not justify giving the populations species rank. Generally, morphological and genetic data diverge in parallel. Two points are especially clear: (1) a species concept is necessary that does not go to the absurdity that one or two qualitative differences among allopatric populations justifies species status, especially on the basis that the separated populations, being thus isolated, have their own unique trajectory in time and space; (2) one needs all the possible data one can obtain to sort out some species problems, starting with detailed anatomical data and ecological observations.

Indeed, the new generation of molecular tools are yielding surprises. Sequence data join the other tools in providing powerful insights into patterns and processes of evolution. However, as mountains of data accumulate, it will become increasingly clear that all the problems with other molecular data sets will become evident with sequence data: convergences, sequences of one molecule (e.g. LrRNA) being uninformative for some groupings of taxa, while showing wild divergences for other taxa, etc. These problems will settle down with the sequencing of whole genes and using genes as characters in phylogenetic analysis. We are a long way from this, as yet, costly and time consuming task. Increased automation of procedures will ease the task.

CONCLUSION

Clearly nature is both capricious and pernicious in how she spins off species and promotes patterns and processes of evolution. It certainly appears this way to a seasoned systematist. The work of discriminating among taxa is clearly complex and multidimensional. There are no rote rules to apply such as stating that species status is achieved when Nei's D equals some artificial value. What is splendid today are the variety of tools that can be applied to solving taxonomic problems. The battery of new molecular tools are especially appreciated and provide the basis for much rigor in testing hypotheses about taxonomic relationships. I have discussed in this lecture the utility of molecular tools for uncovering cryptic species, for studying population genetics, and for studying patterns of speciation and phylogeny.

I hope that I have made clear the point that the fundamental basis of taxonomic discrimination is based on detailed comparative anatomy and cytology. Genetic distances, by themselves, do not serve to define species or higher taxa. There is no universal molecular clock! Further, morphological, allozymic, MtDNA, and DNA sequences may diverge at different rates within the same taxon. Taxa within a clade must be calibrated relative to genetic distance. In studying a situation involving the species-level, it is useful to know if one is involved with an adaptive radiation or a morphostatic radiation; it is useful to know the characters and character-state changes that serve to distinguish species and genera in sister taxa. To add to the want list, a systematist would like to determine the ecological correlates of morphology, the time of taxonomic divergence, and the direction of evolution. Such data require knowing a group on a global basis. Timing and direction may come from paleontological evidence or from geological events. And still, as Giusti and Manganelli (1992) stated so well, a good systematist "..is not above admitting that its exact nature [what is or has occurred] escapes him. . .". Understanding a complex situation in speciation may take years of study.

ACKNOWLEDGMENTS

This presentation, and much of the work presented here was supported by an NIH grant AI 11373 TMP. The graphics were prepared by Susan Trammell. I am indebted to Gary Rosenberg for reading and making useful criticisms of this manuscript.

LITERATURE CITED

Avise, J. C. and C. F. Aquardo. 1982. A comparative summary of genetic distances in the vertebrates. Evolutionary Biology 15:151–158.

Bandoni, S., M. Mulvey, D.Koech and E.Loker. 1990. Genetic structure of Kenyan populations of *Biomphalaria pfeifferi* (Gastropoda: Planorbidae). Journal of Molluscan Studies 56: 383–391.

Barton, N. H. 1989. Founder effect speciation. *In:* D. Otte and J. Endler (eds.) Speciation and Its Consequences. Sinauer Press, Sunderland, MA. pp.256–296.

Benton, M. J. 1988. The nature of an adaptive radiation. Tree 3:127–128.

Bianchi, T. S., G. M. Davis and D. Strayer. 1994. An apparent hybrid zone between freshwater gastropod species *Elimia livescens* and *Elimia virginica* (Mesogastropoda: Pleuroceridae). American Malacological Bulletin (in press).

Buth, D. G. 1984. The application of electrophoretic data in systematics studies. Annual Review of Ecology and Systematics 15: 501–522.

Cavalli-Sforza, L. L. and A. W. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Evolution 21: 550–570.

Chambers, S. M. 1978. An electrophoretically detected sibling species of "*Goniobasis floridensis*". Malacologia 17: 157–162.

Chambers, S. M. 1980. Genetic divergence between populations of *Goniobasis* (Pleuroceridae) occupying different drainage systems. Malacologia 20:63–81.

Cracraft, J. 1989. Speciation and its ontology: The empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. *In:* D. Otte and J. Endler, (eds.) Speciation and its consequences. Sinauer Associates,Sunderland, MA. pp. 28–59.

Davis, G. M. 1969. Reproductive, neural and other anatomical aspects of *Oncomelania minima* (Prosobranchia: Hydrobiidae). Venus 28:1–36.

Davis, G. M. 1979a. The origin and evolution of the gastropod

family Pomatiopsidae, with emphasis on the Mekong River Triculinae. Monograph of The Academy of Natural Sciences No.20:1–120.

Davis, G. M. 1979b. Experimental methods in molluscan systematics. *In:* Fretter, V. and J. Peake (eds.) Pulmonates Vol. 2. Academic Press, New York pp. 99–169.

Davis, G. M. 1980. Snail hosts of Asian *Schistosoma* infecting man: origin and coevolution. *In:* Bruce, J. *et al.* (eds.) The Mekong Schistosome. Malacological Review, Supplement 2:195–238.

Davis, G. M. 1981. Different modes of evolution and adaptive radiation in the Pomatiopsidae (Prosobranchia: Mesogastropoda). Malacologia 21:209–262.

Davis, G. M. 1983. Relative roles of molecular genetics, anatomy, morphometrics, and ecology in assessing relationships among North American Unionidae (Bivalvia). *In:* Oxford, G. S. and D. Rollinson (eds.) Protein Polymorphism: Adaptive and Taxonomic Significance. Systematics Association Special Volume No.24, Academic Press, New York pp. 193–221.

Davis, G. M. 1984. Genetic relationships among some North American Unionidae (Bivalvia): Sibling species, convergence, and cladistic relationships. Malacologia 25:629–648.

Davis, G. M. 1992. Evolution of prosobranch snails transmitting Asian *Schistosoma*; coevolution with *Schistosoma*: A review. Progress in Clinical Parasitology 3:145–204.

Davis, G. M. (ed.) Systematics and Public Health:A contribution to the Systematics Agenda 2000 (in press).

Davis, G. M., C. E. Chen and S. H. Yu. 1994a. Unique morphological innovation and population variation in *Gammatricula songi*, a new species of Triculinae from China (Gastropoda: Rissoacea). Proceedings of the Academy of Natural Sciences of Philadelphia 145:107–145.

Davis, G. M., C. E. Chen, X. P. Zeng, S. H. Yu and L.Li. 1994b. Molecular genetic and anatomical relationships among pomatiopsid (Gastropoda: Prosobranchia) genera from southern China. Proceedings of The Academy of Natural Sciences of Philadelphia 145:191–207.

Davis, G. M., V. Forbes and G. Lopez. 1988. Species status of northeastern American *Hydrobia* (Gastropoda:Prosobranchia): Ecology, morphology and molecular genetics. Proceedings of The Academy of Natural Sciences of Philadelphia 140:191–246.

Davis, G. M. and S. L. H. Fuller. 1981. Genetic relationships among recent Unionacea (Bivalvia) of North America. Malacologia 20:217–253.

Davis, G. M., W. H. Heard, S. L. H. Fuller and C. Hesterman. 1981. Molecular genetics and speciation in *Elliptio* and its relationships to other taxa of North American Unionidae (Bivalvia). Biological Journal of the Linnean Society 15: 131–150.

Davis, G. M. and Z. B. Kang. 1990. The genus *Wuconchona* of China (Gastropoda: Pomatiopsidae: Triculinae): Anatomy, systematics, cladistics and transmission of *Schistosoma*. Proceedings of the Academy of Natural Sciences of Philadelphia 142:119–142.

Davis, G. M., Y. H. Kuo, K. E. Hoagland, P. L. Chen, H. M. Yang and D. J. Chen. 1984. *Kunmingia*, a new genus of Triculinae (Gastropoda: Pomatiopsidae) from China: Phenetic and cladistic relationships. Proceedings of the Academy of Natural Sciences of Philadelphia 136:165–193.

Davis, G. M., M. McKee and G. Lopez. 1989. The identity of *Hydrobia truncata* (Gastropoda: Hydrobiinae): comparative anatomy, molecular genetics, ecology. Proceedings of The Academy of Natural Sciences of Philadelphia 141:333–359.

Dillon, R. T. 1984. Geographic distance, environmental difference, and divergence between isolated populations. Systematic Zoology 33:69–82.

Dillon, R. T. 1988. Evolution from transplants between genetically distinct populations of freshwater snails. Genetica 76:111–119.

Dillon, R. T. and G. M. Davis. 1980. The *Goniobasis* of southern Virginia and northern North Carolina: genetic and shell morphometric relationships. Malacologia 20:83–98.

Emberton, K. C., G. S. Kuncio, G. M. Davis, S. M. Phillips, K. M. Monderwicz and Y. H. Guo. 1990. Comparison of recent classifications of stylommatophoran land-snail families, and evaluation of large-ribosomal-RNA sequencing for their phylogenetics. Malacologia 31:327–352.

Farris, J. 1988. HENNIG 86,Version 1.5. Computer Software. Stonybrook, NY.

Gillespie, J. H. 1984. The molecular clock may be an episodic clock. Proceedings of the National Academy of Science. U.S.A. 81:8009–8013.

Gillespie, J. H. 1986a. Rates of molecular evolution. Annual Review of Ecology and Systematics 17:637–665.

Gillespie, J. H. 1986b. Variability of evolutionary rates of DNA. Genetics 113:1077–1091.

Gillespie, J. H. 1987. Molecular evolution and the neutral allele theory. Oxford Surveys Evolutionary Biology 4:10–37.

Gittenberger, E. 1991. What about non-adaptive radiation? Biological Journal of the Linnean Society 43:263–272.

Giusti, F. and G. Manganelli. 1992. The problem of the species in malacology after clear evidence of the limits of morphological systematics. *In:* E. Gittenberger, E. and J. Goud (eds.). Proceedings of the Ninth International Malacological Congress, Edinburgh, 1986. Unitas Malacologica, Leiden. pp. 153–172.

Harrison, R. G. 1991. Molecular changes at speciation. Annual Review of Ecology and Systematics 22:281–308.

Hewitt, G. 1988. Hybrid zones - Natural laboratories for evolutionary studies. Trends in Ecology and Evolution 3:158–167.

Hillis, D. M. and C. Moritz. 1990. An overview of applications of molecular systematics. *In:* Hillis, D. and C. Moritz (eds.) Molecular Systematics. Sinauer Associates, Inc. Sunderland, MA pp. 502- 515.

Hoagland, K. E. and G. M. Davis. 1987. The succineid snail fauna of Chittenango Falls, New York: Taxonomic status with comparisons to other relevant taxa. Proceedings of The Academy of Natural Sciences of Philadelphia 139: 465–526.

Hornbach, D. J., M. J. McLeod, S. I. Guttman and S. K. Seilkop. 1980. Genetic and morphological variation in the freshwater clam, *Sphaerium* (Bivalvia: Sphaeriidae). Journal of Molluscan Studies 46:158–170.

Johnson, M. S., O. C. Stine and J. Murray. 1984. Reproductive compatibility despite large-scale genetic divergence in *Cepaea nemoralis*. Heredity 53:655–665.

Johnson, R. I. 1970. The systematics and zoogeography of the Unionidae (Mollusca: Bivalvia) of the southern Atlantic slope region. Bulletin of the Museum of Comparative Zoology at Harvard. 140:263–450.

Kat, P. W. 1983a. Genetic and morphological divergence among nominal species of North American *Anodonta* (Bivalvia; Unionidae). Malacologia 23:361–374.

Kat, P. W. 1983b. Genetics, morphological divergence, and speciation among *Lampsilis*. Journal of Molluscan Studies 49:133–145.

Kirpatrick, M. and R. K. Selander. 1979. Genetics of speciation in lake whitefish in the Allegash Basin. Evolution 33: 478–485.

Landergren, U., R. Kaiser, C. T. Caskey and L. Hood. 1988. DNA diagnostics - molecular techniques and automation. Science 242:229–237.

Li, W. H., M. Tanimura and P. M. Sharp. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. Journal of Molecular Evolution 25:330–342.

Mayr, E. 1963. Animal Species and Evolution. Harvard University Press, Cambridge, MA.

McCracken, G. F. and R. K. Selander. 1980. Self-fertilization and monogenic strains in natural populations of terrestrial slugs. Proceedings of the National Academy of Sciences 77:684–688

Mulvey, M., M. C. Newman and D. S. Woodruff. 1988. Genetic differentiation among West Indian populations of the schistosome-transmitting snail *Biomphalaria glabrata*. Malacologia 29:309–317.

Murray, J., O. C. Stine and M. S. Johnson. 1991. The evolution of mitochondrial DNA in *Partula*. Heredity 66:93–104.

Nei, M. 1972. Genetic distance between populations. American Naturalist 106:283–292.

Nei, M. and D. Graur. 1984. Extent of protein polymorphism and the neutral mutation theory. *In:* Hecht, M., B. Wallace and G. Prance (eds.) Evolutionary Biology 17:73–118.

Nelson, G. 1989. Species and taxa: Systematics and Evolution. *In:* Otte, D. and J. Endler (eds.) Speciation and its Consequences. Sinauer Associates, Inc. Sunderland, MA pp: 60–81.

Nelson, G. and N. Platnick. 1981. Systematics and Biogeography: Cladistics and Vicariance. Columbia University Press, New York. 567 pp.

Nevo, E. 1991. Evolutionary theory and process of speciation and adaptive radiation in subterranean mole rats, *Spalax ehrenbergi* superspecies, in Israel. *In:* M. Hecht, B. Wallace and R. Macintyre (eds) Evolutionary Biology 25:1–125.

Nevo, E. and A. Beiles. 1988. Genetic parallelism of protein polymorphism in nature: ecological test of neutral theory of molecular evolution. Biological Journal of the Linnean Society 35:229–245.

Osborn, H. F. 1918. The Origin and Evolution of Life. Scribner's Sons, New York. 322 pp.

Patterson, H. E. H. 1985. The recognition concept of species. *In:* Vrba, E. S. (ed.) Species and Speciation. Transvaal Museum Monograph No.4:21–29.

Rosen, D. 1978. Vicariant patterns and historical explanation in biogeography. Systematic Zoology 27:159–188.

Rosenberg, G. 1989. Phylogeny and evolution of terrestriality of the Atlantic Truncatellidae. Ph.D. Dissertation. Harvard University.

Rosenberg, G., G. S. Kuncio, G. M. Davis and M. G. Harasewych. 1993. Ribosomal RNA phylogeny of selected gastropod and unionacean bivalve mollusks. *In:* Harasewych, M. G. and S. Tillier (eds.) Molecular Techniques and Molluscan Phylogeny: Proceedings of a symposium held at the Eleventh International Malacological Congress, Siena, 1992. The Nautilus Supplement 2:111–121 (This volume).

Selander, R. K. and D. W. Kaufman. 1973. Self-fertilization and genetic population structure in a colonizing landsnail. Proceedings of the National Academy of Sciences. 70:1186-1190.

Selander, R. K. and H. Ochman. 1983. The genetic structure of Pulmonates as illustrated by Molluscs. *In:* Isozymes: Current Topics in Biological and Medical Research 10:93–123. Alan R. Liss Inc.,N.Y.

Sene, F. M. and H. L. Carson. 1977. Genetic variation in Hawaiian *Drosophila*. IV Allozymic similarity between *D.silvestris* and *D. heteroneura* from the island of Hawaii. Genetics 86:187–198.

Skibinski, D. and R. D. Ward. 1982. Correlations between heterozygosity and evolutionary rate of proteins. Nature 298:490–492.

Stanley, S. M. 1979. Macroevolution. W. H. Freeman, San Francisco, CA 364 pp.

Stine, O. C. 1989. *Cepaea nemoralis* from Lexington, Virginia: The isolation and characterization of their mitochondrial DNA, the implications for their origin and climatic selection. Malacologia 30:305–315.

Swofford, D. L. 1983. PAUP: Phylogenetic Analysis Using Parsimony. Computer Software. Illinois Natural History Survey, Champaign,Illinois.

Swofford, D. L. and G. J. Olsen. 1990. Phylogeny Reconstruction. *In:*Hillis, D. and C. Moritz (eds.) Molecular Systematics. Sinauer Ass. Sunderland, MA pp. 411–501.

Templeton, A. R. 1989. The meaning of species and speciation: a genetic perspective. *In:* D. Otte and J. Endler (eds.) Speciation and its Consequences. Sinauer Associates, Inc., Sunderland, MA pp. 3–27.

Templeton, A. R. 1980. Modes of speciation and inferences based on genetic distances. Evolution 34:719–729.

Templeton, A. R. 1981. Mechanisms of speciation - a population genetic approach. Annual Review of Ecology and Systematics 12:23–84.

Vainola, R. and M. M. Huilsom. 1991. Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations (Mytilidae: Mollusca). Biological Journal of the Linnean Society 43: 127–148.

Wiley, E. O. 1981. Phylogenetics. J. Wiley, New York.

Woodruff, D. S. 1989. Genetic anomalies associated with *Cerion* hybrid zones: the origin and maintenance of new electromorphic variants called hybrizymes. Biological Journal of the Linnean Society 36:281–294.

Woodruff, D., K. C. Staub, E. S. Upatham, V. Viyant and H. C. Yuan. 1988. Genetic variation in *Oncomelania hupensis*: *Schistosoma japonicum* transmitting snails in China and the Philippines are distinct species. Malacologia 29:347–361.

Wright, S. 1978. Evolution and the Genetics of Populations. Vol.4. The University of Chicago Press. 580 pp.

Zimmer, E. A., S. L. Martin, S. M. Beverly, Y. W. Kan and A. C. Wilson. 1980. Rapid duplication and loss of genes coding for the chains of hemoglobin. Proceedings of the National Academy of Science, USA. 77:2158–2162.

The Evolutionary Consequences of Restrictions on Gene Flow: Examples from Hydrobiid Snails

D. J. Colgan
W. F. Ponder
The Australian Museum
P.O. Box A285
Sydney South, Australia 2000

ABSTRACT

The evolutionary consequences of restrictions on gene flow are discussed in relation to the population genetic structure and speciation of four Australian hydrobiid snail faunas. The studied faunas comprise: (1) the brackish water genus *Tatea*; (2) species of *Fluvidona* in freshwater streams at Wilsons Promontory; (3) species of *Fonscochlea* and *Trochidrobia* in artesian springs near Lake Eyre in South Australia; and (4) species of an undescribed genus at Dalhousie Springs in northern South Australia, another arid zone artesian spring complex. Gene flow in these hydrobiids is very variable. It is high in *Tatea*, relatively high at Dalhousie Springs and extremely low in Wilsons Promontory and the Lake Eyre springs. Levels in the latter two faunas are similar despite a great disparity in geographic area. In these four faunas, the detail of gene flow patterns is complex, emphasising the dependence of population structure on the interaction of current and historical factors. This is illustrated by speciation patterns, the numbers of species and their distributions usually correlating well with observed levels of gene flow. However, there are examples, among species groups with comparable, but very low gene flow, in which some taxa have undergone speciation yet others have not.

The data were analysed using F-statistics and the private allele frequency approaches. Whilst the qualitative conclusions from the two approaches were generally similar, the exceptions usually indicated (on biogeographic grounds), that the F-statistics approach is the more reliable estimator of gene flow. The private alleles approach is dependent on a coincidence of the scale of sampling with the biological scale of population subdivision. Intensive sampling schemes, as utilised in our studies, tend to find even rare alleles in more than one population even though they may be quite restricted in geographical distribution. An analytical method for treating conditional allelic frequencies would not be as sensitive to this problem as the private alleles approach.

Key words: Gene flow, snails, freshwater, isozymes, Hydrobiidae, evolution, population structure.

INTRODUCTION

The evolutionary fate of populations is largely determined by two sets of factors. The first may be characterised as "coping factors"—those determining the survival of a population. This set includes external pressures, such as predation, parasitism and disease, extremes of, or changes in, climate, competition from other species or resource diminution. It also includes endogenous properties such as the amount of local inbreeding or the capacity of the breeding system to engender genetic recombination. The second set may be characterised as "isolating factors"—those factors which affect the evolutionary history of a population relative to other, originally con-specific, populations. In general, this history depends on how effective gene flow is in overcoming differentiation inevitably arising from genetic drift or responses to local selection. Conversely, speciation processes are contingent on an "evolutionarily sufficient" restriction of gene flow between populations that have successfully accommodated the first set of factors. These two sets of factors are, however, also inter-related. For instance, it may be only through the introduction of a novel gene from another area that a population is enabled to withstand a climatic change.

Many studies have shown that speciation is directly related to reductions in gene flow (reviewed by Grant, 1980; Porter, 1990). It is difficult, however, if not impossible, to predict what restrictions on gene flow over what period of time constitute evolutionarily significant barriers. The importance of the evolutionary consequences of restriction on gene flow is such that its estimation remains a goal of many experimental studies (*e.g.* Skibinski *et al.*, 1983; Waples, 1987; Johnson *et al.*, 1988; Mitton *et al.*, 1989; Arter, 1990; Porter, 1990; Preziosi & Fairbairn, 1992). There is also continuing interest in the development of mathematical models for the analysis of gene flow (*e.g.* Slatkin, 1985a; Barton & Slatkin, 1986; Slatkin & Barton, 1989) and/or the the genetical subdivision of species (Cavalli-Sforza & Feldman, 1990).

We have been investigating the hydrobiid gastropod faunas of a variety of habitats to characterise their taxonomic and population genetic structure. We were particularly interested in how biological and environmental differences between these faunas are reflected in the

degree of genetic divergence and the amount of genetic exchange between populations, and in the evolutionary consequences of those differences. *A priori*, factors such as the biological ability to disperse, the geographic scale of the system, weather patterns, topography and the accessibility and suitability of the habitat for potential biological dispersal agents might all be expected to play a part in determining levels of gene flow.

Gene Flow

In this paper we define gene flow as the genetically-effective transmission of alleles between extant discrete populations and between various parts of the range within species in which population boundaries cannot be discerned or do not occur. We do not regard re-colonization after local extinction as an example of gene flow (agreeing with Endler, 1973; Grant, 1980;—but contrast Slatkin 1985b). Re-colonization, or original colonization simply increases the number of populations of a species, other processes being required for phylogenetic consequences. Generally, continued gene exchange will homogenize original and derived populations. Significant differentiation requires persistent marked reductions in gene flow.

Any estimate of gene flow made on the basis of allelic frequencies confounds factors operating during two distinct phases of the differentiation process. Firstly, the establishment of a population implies a sampling process which may cause differences in frequency arrays (Carson & Templeton, 1984; Wool, 1987). Secondly, the differences reflect subsequent patterns of gene flow and differential selection as well as any current trends. In the discussion below, we usually make the assumption that the comparisons between estimates of gene flow in different biological situations reflect differences in only one of these two confounded factors in any given case. This may not always be an accurate description of biological reality. High allelic frequency differences may result from a divergent initial founder effect or from a subsequent reduction in gene flow, differential selection or any combination of these three factors. Direct methods of estimating gene flow, requiring observation of the mating success and/or fertility of known immigrant individuals, largely overcome these problems. The labour and practical difficulies involved in making such estimates is such, however, that indirect methods are usually pursued (Slatkin, 1985b; Johnson *et al.*, 1988).

The Measurement of Gene Flow

Both methods of estimating gene flow which are used here measure the parameter Nm, where N is the (effective) size of each sub-population and m is the probability that a gamete in the offspring generation is an immigrant to the sub-population where it occurs. Three principal models of population structure have been developed as mathematical abstractions to provide a theoretical framework for measuring Nm:

(A) The Island model (Wright, 1931), in which each of an infinite number of discrete sub-populations receives migrants at random from other sub-populations. The geographic distance between populations does not affect the rate of gene flow between them.

(B) The Stepping Stone model (Kimura & Weiss, 1964), in which gene flow occurs between a population and its immediate neighbours in one or two dimensional geographic arrays. Gene flow does not occur directly between two populations which are not immediate neighbours.

(C) Models in which the population is considered to be continuously distributed in one or two dimensions, with the degree of genetic differentiation of individuals separated by a given distance determined by the levels of gene flow.

The accuracy of these models' approximation to population structure will vary. If the studied species has high vagility, the requirement of the Island model that each sub-population exchanges migrants with all others will be a more accurate approximation than if the vagility is low. Conversely, the Stepping Stone model's restriction on migration between populations which are not near neighbours is more likely to be accurate if the studied species has low vagility.

F-Statistics

The overall inbreeding coefficient can be partitioned into components reflecting non-random breeding (F_{IS}) and the effects of between sub-population differentiation (F_{ST}) (Wright, 1951). Under the infinite island model of population subdivision, if the migration rate is small, then (Wright, 1951):

$$F_{ST} \approx (1 + 4Nm)^{-1}$$

A variety, indeed almost a plethora, of alternative methods for the calculation of quantities very similar or identical to F_{ST} have been suggested (Wright, 1951, 1978; Nei, 1973; Nei & Chesser, 1983; Cockerham, 1969; Weir & Cockerham, 1984). Many of these were developed in response to complications of the original two-allele per locus situation studied by Wright. It can be shown that these are usually encompassed by natural extensions of Wright's approach. Others attempt to take account of relaxation of the simplifying assumptions (negligible selection, mutation, *etc.*) made in Wright's analyses. The various methods have been widely reviewed (e.g. Chakraborty & Leimar, 1987; Weir, 1990) and the differences between them shown, generally, to be of second-order significance. Moreover, both analytical (Slatkin, 1985b) and simulation (Slatkin & Barton, 1989) studies tend to emphasise the qualitative similarities between F_{ST} variables defined under either the Island or Stepping Stone Models. Where estimates of gene flow given below are based on F_{ST}, this will be indicated by $Nm(F_{ST})$.

Conditional Allelic Frequencies

There are two main methods of analysing gene flow using the approaches of Slatkin (1981, 1985a). The first requires that the "occupancy rate" for an allele be determined. This is the number of sub-populations in which the allele is found, divided by the total number of sub-populations. The conditional average frequency of the allele is the average of its frequencies in those sub-populations where it actually occurs. Levels of gene flow between sub-populations are visualised by graphing the conditional average frequency of each allele against occupancy rate. Such representations can be useful in comparison of the levels of gene flow in different taxa (*e.g.* Govindaraju, 1989) but, in the absence of an analytical theory, can be used to provide numerical estimates only by making analogies with the results of computer simulations (Johnson *et al.*, 1988). In the second method, attention is restricted to "private alleles", *i.e.* those found in only one sub-population. Slatkin's simulations (1985a) found that, in both Island and Stepping Stone Models, the conditional average frequency of private alleles ($\bar{p}(1)$) is approximately linearly related to the migration rate by the expression:

$$\log_{10}(\bar{p}(1)) \approx a\log_{10}(Nm) + b$$

where a and b take values dependent on the number of individuals sampled from each sub-population. One claimed advantage of this approach is that its estimates of migration rates are theoretically only slightly dependent on mutation or the many types of selection which might operate (Barton & Slatkin, 1986; Slatkin & Barton, 1989). Estimates of gene flow based on the frequency of private alleles given below are designated as $Nm(\bar{p}(1))$.

The Family Hydrobiidae

Small prosobranch snails of the world-wide family Hydrobiidae are the most diverse freshwater gastropods, with nearly 400 generic names currently in use (Kabat & Hershler, 1993). Commonly, in Australia, freshwater Hydrobiidae occupy small streams or springs. The populations in these isolated or semi-isolated habitats show varying degrees of differentiation because of an apparent inability to disperse readily. A low level of dispersal may be possible, for example by birds or even flying insects (Rees, 1965; Boeters, 1979, 1982).

Other genera inhabit brackish or estuarine waters. Genetic data have been used to test hypotheses based on morphological criteria in such taxa (Lassen, 1979; Davis *et al.*, 1988, 1989; Ponder & Clark, 1988; Ponder *et al.*, 1991). They tend to have large geographic ranges, partly because some have a planktonic marine larval phase, but also because they live in tidal marshland habitats where they are potentially readily transported by birds.

In the remainder of this paper, we will concentrate on our recent investigations of three freshwater hydrobiid radiations at Wilsons Promontory, in the Lake Eyre supergroup of the South Australian Mound Springs and in the Dalhousie Springs complex at the north of South Australia. For comparative purposes, we will often refer to our studies of the brackish water (usually estuarine) genus *Tatea* (Ponder *et al.*, 1991). The fauna of the Lake Eyre spring supergroup has been formally described (Ponder *et al.*, 1989) and that of Dalhousie Springs has been briefly reported on with respect to shell morphology (Ponder, 1989). The Wilsons Promontory study will be described in detail in a forthcoming publication (Ponder *et al.*, 1994).

MATERIALS AND METHODS

Summary information regarding collecting sites, *etc.* can be found in Appendix 2. More detail is provided for *Tatea* in Ponder *et al.* (1991), the Wilsons Promontory *Fluvidona* in Ponder *et al.* (1994), the Lake Eyre springs in Ponder *et al.* (1989) and on Dalhousie Springs in Zeidler and Ponder (1989). Genotypic data are available from the senior author. Data for *Tatea* and *Fluvidona* on allozymic frequencies, observed heterozygosities and the various environmental parameters which were measured are given in Ponder *et al.* (1991, 1994). Similar data for the other faunas will be presented separately for each system.

Standard methods for cellulose acetate electrophoresis were used (Hebert & Beaton, 1989, Ponder *et al.*, 1991). Individual snails were homogenized with 10–30 μl (mean 20 μl) of buffer, providing enough sample for up to 12 gels. Because of their small size, it was not possible to examine each snail for all enzymes. Where more than one locus is shown below as encoding the same enzyme, each was designated numerically in order of decreasing mobility. Allozymes identified for each locus are designated in the same way. The enzymes scored for each species, or species grouping, together with abbreviations used, Enzyme Commission Numbers, and number of presumptive loci are listed in Appendix 1. There were some differences between taxa in the number of loci that were electrophoretically interpretable. These are specified in Appendix 1. The computer packages BIOSYS–1 (Swofford & Selander, 1981), NTSYS (Rohlf, 1990) and PHYLIP, version 3–4 (Felsenstein, 1989) were used to assist analysis.

All taxonomic groupings treated here were initially analysed without assuming any hierarchical structure of the populations. F_{ST}, conditional allozymic frequencies and private allelic frequencies were calculated. Populations were then clustered in hierarchies, as described below and as detailed in Appendix 2. Components of overall genetic differentiation were obtained for this clustering using the WRIGHT78 step of BIOSYS. Allozymic frequencies in taxonomic units at each intermediate level of the clustering were calculated after pooling the data from the sub-units included in the same group. The pooled data were used to estimate F_{ST} values and conditional (and private) allelic frequencies for units at this intermediate level. This approach has a statistical tendency to reduce the variance in gene frequencies (and hence

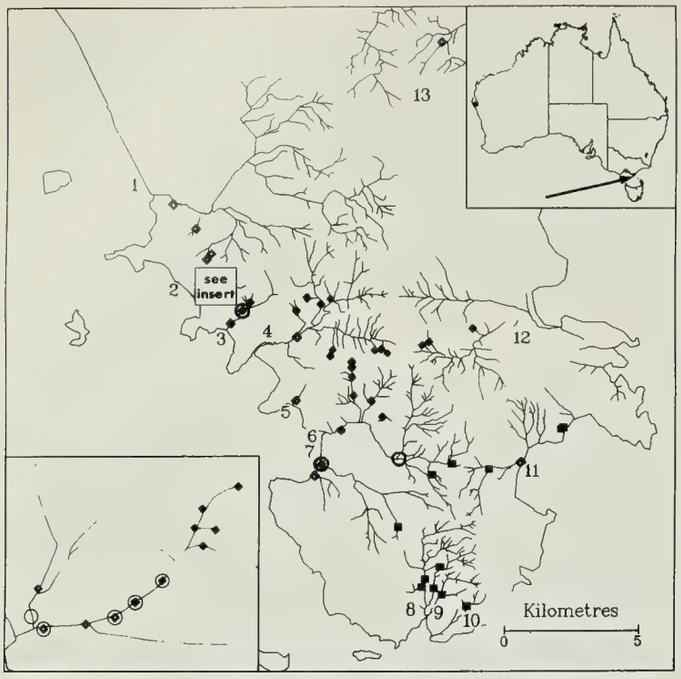


Figure 1. Map of Wilsons Promontory showing major drainages and the distribution of *Fluvidona* species. The area covered is shown by the dot at the head of the arrow on the inset map of Australia. The other inset shows a detail of Whisky Creek. Locations of straight-sided snails are shown by ○, the MPI 3 convex by ■, MPI 4 by ◆ and MPI 5 by ◇. Drainages: (1) Darby River; (2) Whisky Creek; (3) Squeaky Creek; (4) Tidal River; (5) Titania Creek; (6) Growler Creek; (7) Frasers Creek; (8) Roaring Meg; (9) Picnic Creek; (10) First Bridge Creek; (11) Freshwater Creek; (12) Blackfish Creek; and (13) Chinamans Creek.

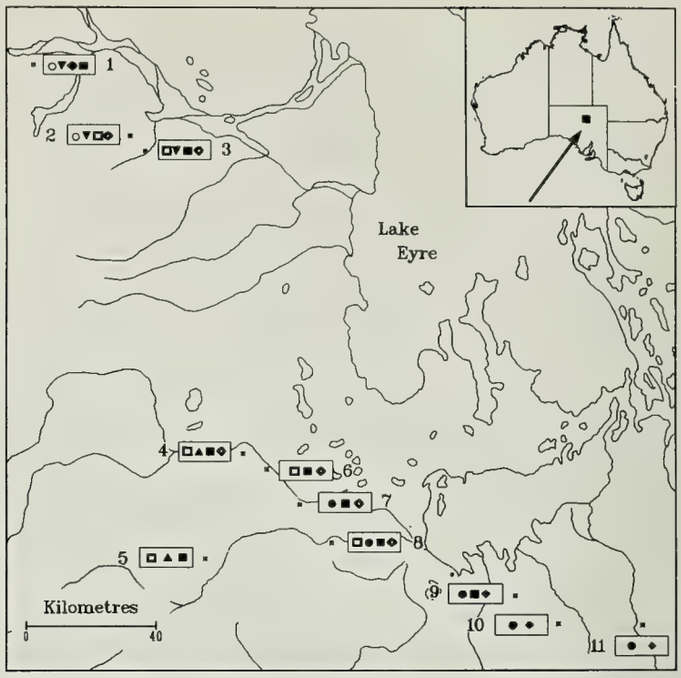


Figure 2. Map of the Lake Eyre mound springs showing the distribution of *Fonscochlea* and *Trochidrobia* species. The area covered is indicated on the inset of Australia. The site of a spring group is indicated by an "x". Presence of a species in the group is indicated by ◆ for *F. accepta*, ◇ for *F. aquatica*, ■ for *F. zeidleri*, ▲ for *F. billakalina*, ▼ for *F. variabilis*, ● for *T. punicea*, ○ for *T. minuta* and □ for *T. smithi*. The spring groups are: (1) Freeling; (2) Outside; (3) Twelve Mile; (4) Strangways; (5) Billakalina; (6) Beresford/Warburton; (7) Coward/Jersey/Elisabeth/Kewson; (8) Blanche Cup; (9) Hermit Hill; (10) Davenport; and (11) Welcome.

increase the estimate of *Nm*) to a degree dependent on levels of variation between samples pooled into the same unit. If the variation is merely a sampling artefact then the procedure will increase accuracy of *Nm* estimation. But if the variation is due to biological subdivision of the populations, then the estimate should be regarded more as an upper limit on the degree of gene flow. *A priori*, it is not possible to decide which of these two alternatives is correct as we do not know what constitutes an effectively panmictic unit in these hydrobiids. A second effect of the pooling procedure is that alleles which are found in more than one population and hence not "private" in the original subdivision, may be regarded as private at higher clustering levels if they are there restricted to only one unit. Again, this reflects the uncertainty about the biological structure of the population. Pooling data may not always resolve this uncertainty but patterns in such analyses will usually be informative about population structure to at least some extent.

The sample sizes used in the studies varied from locus to locus and from population to population. The average sample per locus is shown in Appendix 2. When *Nm* was estimated from the conditional frequencies of private alleles, the parameter values for a sample size of 25 were taken from Barton and Slatkin (1986). Alternative parameter values did not, however, significantly affect numerical estimates of gene flow.

THE STUDY AREAS

Wilsons Promontory

Wilsons Promontory (Figure 1), the southern-most part of the Australian mainland (39°S, 146°28'W), consists of granite hills up to 754 m with many permanent streams and rivers fed by high rainfall (>1000 mm per year). Its geological and climatological setting are summarised by Wallis (1988) and Schmidt and Thornton (1992). The hydrobiid fauna (genus *Fluvidona*) of the Promontory comprises two endemic morphologically-recognisable species, the shells of one with straighter whorl outlines and one with more convex whorls. The latter morphospecies is divisible into three genetic species by very nearly fixed sympatric differences in the MPI phenotype, referred to as the MPI 3, MPI 4 and MPI 5 genetic species.

For hierarchical analyses of population structure, sites were grouped within streams, streams within catchments and catchments within species, providing two intermediate levels (streams, catchments) in an analysis. Populations in which hybrids were seen were ignored in the

analyses examining gene flow within genetic species (see Appendix 2).

South Australian Mound Springs

The springs, fed from the Great Artesian Basin of Australia, are of considerable limnological and conservation significance to the very arid area in which they occur (Ponder, 1986; Harris, 1993). The springs generally lie on the fringes of the Basin, where the aquifers abut impervious rock or lie near the surface. The Lake Eyre Supergroup, the most extensive group of springs associated with the Great Artesian Basin, extends about 400 km between Marree and Oodnadatta and provides virtually the only permanent water in the area. The area with springs is only rarely more than 20 km wide, so that the supergroup conforms quite well to a one dimensional, discontinuous model. The geological history of these springs is not well known. Estimates of the ages of some large (extinct) mounds range from late Miocene to Recent. These are probably at least Pleistocene (Wopfner & Twidale, 1976; Williams & Holmes, 1978; Thompson & Barnett, 1985) but the springs have probably been in the area much longer. The taxa presently inhabiting the springs may be relicts of more widespread forms from a generally wetter period in the Neogene or may represent faunas associated with these artesian springs through much of the Tertiary.

The predominant drainage pattern in the Lake Eyre basin is at right angles to the line of springs. Hence the transport of snails between spring groups by floods would be unlikely—although such transport could occur within spring groups. Spring nomenclature and grouping used in this paper follows Ponder *et al.* (1989). The hydrobiid fauna consists of two endemic genera, *Fonscochlea* (five species) and *Trochidrobia* (four species) (Ponder *et al.*, 1989)

For hierarchical analyses, springs were compared within spring groups. The groups were then collected into "clusters" (see Figure 2, Appendix 2), the Southern cluster comprising springs between Welcome Springs and Hermit Hill, the Middle cluster comprising springs between the Blanche Cup complex and Strangways and the Northern comprising Outside, Twelve Mile and Freeling Springs. If species were found in more than one of these clusters, a second intermediate level was included in the hierarchy. This could not, however, be done for all species, *Fonscochlea accepta*, for instance being found only in the Southern cluster. The Southern cluster was divided into three groups: (1) Welcome Springs; (2) Davenport Springs; and (3) the Hermit Hill springs. The Middle cluster was divided into five groups: (1) Blanche Cup springs, (2) the group consisting of Coward, Kewson, Elizabeth and Jersey springs, (3) Billakalina; (4) Beresford and Warburton Springs; and (5) Strangways Springs. The Northern cluster was divided into two groups: (1) Twelve Mile and Outside springs; and (2) Freeling Springs.

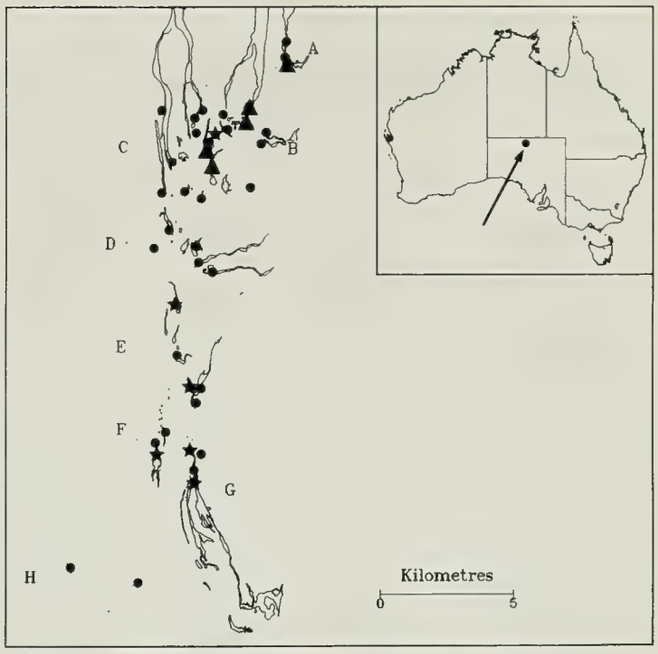


Figure 3. Map of the Dalhousie Springs showing the distribution of the globular (▲), pupiform (●) and *Fluvidona*-like snails (★). Spring groups are identified by letters near dashed boundaries.

Dalhousie Springs

This complex is another large group of arid zone springs in northern South Australia associated with the Great Artesian Basin (Figure 3). Aspects of its geology and biology are surveyed in a number of papers in Zeidler and Ponder (1989). The many springs in the complex occupy an area of about 70 km^2. They range in size from small nascent or senescent seeps, to actively flowing and, at the upper end of the size scale, to outlets (of about 140 L/sec) which feed pools of 50 m or more in width with outflow channels supporting wetland vegetation for up to 15 km (Smith, 1989). Their combined discharge accounts for 90–95% of the total produced by all South Australian artesian springs (Smith, 1989), and 41% of the overall output from Great Artesian Basin springs (Habermehl, 1982). They are well separated from springs of the Lake Eyre supergroup, the northernmost population of species from that region (*F. zeidleri*) being 140 km away. These springs are likely to be early Pleistocene in age (Krieg, 1989). Minor local overflow due to rare heavy rain may facilitate interspring transport. Major flooding is unlikely (Kotwicki, 1989).

Spring nomenclature and groupings used in this paper follow Zeidler and Ponder (1989), except as specified below. Eight main groups of springs, designated A to H are recognised (Figure 3, Appendix 2). C is divided into four sub-groups, and D into two. Herein, group H will be treated as comprising two groups, because H3 is well separated from H1. We have also split E into two subgroups containing, respectively, (1) E5 and E1 and (2)

Table 1. Estimates of *Nm* derived from the average frequency of private alleles or from the average F_{ST} at various clustering levels. The overall estimates assume no population hierarchy. The next two columns are estimates from data pooling within the first intermediate hierarchical level and the final two columns are for pooling within the second hierarchical level (where applicable). The upper figure in each cell is the observed value of the variable. The lower figure is the value of *Nm* calculated from the observation.

	Overall		Level 1 pooling		Level 2 pooling	
Species	F_{ST}	$\bar{p}(1)$	F_{ST}	$\bar{p}(1)$	F_{ST}	$\bar{p}(1)$
Fluvidona (straight-sided)	0.130	0.444	0.120	0.375	0.108	0.348
	0.776	0.313	0.491	0.417	0.589	0.468
Fluvidona (convex) MPI 3	0.086	0.535	0.110	0.523		
	0.872	0.217	0.570	0.228		
MPI 4	0.075	0.681	0.199	0.628	0.081	0.462
	1.104	0.117	0.205	0.148	0.967	0.291
MPI 5	0.161	0.240	0.091	0.282	0.073	0.254
	0.295	0.791	0.791	0.637	1.157	0.734
Fonscochlea accepta	0.116	0.207	0.028	0.110		
	0.525	0.958	6.036	2.023		
F. aquatica	0.110	0.781	0.128	0.756	0.149	0.412
	0.570	0.070	0.439	0.081	0.338	0.357
F. zeidleri	0.360	0.728	0.289	0.750	0.155	0.625
	0.074	0.093	0.108	0.083	0.316	0.150
F. variabilis	0.196	0.792	0.300	0.783	0.173	0.688
	0.211	0.066	0.101	0.069	0.261	0.113
F. billakalina	0.044	0.263	0.226	0.300		
	2.769	0.700	0.165	0.583		
Trochidrobia punicea	0.143	0.600	0.098	0.580	0.160	0.470
	0.363	0.167	0.696	0.175	0.299	0.282
T. smithi	0.426	0.730	0.265	0.584	0.224	0.346
	0.055	0.092	0.125	0.178	0.167	0.472
T. minuta	0.086	0.375	0.244	0.429		
	0.872	0.417	0.144	0.381		
Dalhousie (globular)	0.159	0.321	0.038	0.105	0.028	0.082
	0.302	0.529	3.565	2.131	6.036	2.799
Dalhousie (pupiform)	0.180	0.413	0.015	0.296	0.014	0.228
	0.244	0.355	17.706	0.595	19.942	0.846
Dalhousie (*Fluvidona*-like)	0.296	0.670	0.300	0.492		
	0.104	0.123	0.101	0.258		

E2, E7 and E8. Generally, for hierarchical analyses, springs were clustered into sub-groups, sub-groups into groups and groups into the species. Three hydrobiid species were recognised in our genetic studies. The globular and pupiform species belong to an undescribed endemic genus, the third species (also endemic) being tentatively included in the widespread genus *Fluvidona*. Separate analyses were performed for each species.

RESULTS

The levels of gene flow in the four *Fluvidona* taxa from Wilsons Promontory are extremely low as shown by the overall F-statistics in Table 1. The smallest F_{ST} value is for the convex MPI 5 genetic species. At 0.240, this is, however, near the higher end of the range previously found for gastropods over comparable geographic scales (Gould & Woodruff, 1986, 1990; Johnson *et al.*, 1988). The overall F_{ST} values for the straight-sided *Fluvidona* species and the convex MPI 3 and MPI 4 genetic species are very high. The levels of migration suggested by these values range down to 0.117 for the MPI 4 genetic species, implying that the fraction of a deme which is replaced by immigrants each generation ($m = 0.117/N$) is very small. The estimates of $Nm(F_{ST})$ values based on the pooling of data from individual samples may be complicated by the likelihood that the pooled data do not represent single populations. The trends in the estimates are, however, very similar to those based on single populations. Those for the MPI 4 genetic species are higher than those for other *Fluvidona* taxa, but still suggest that

Table 2. Variance components in the hierarchical *F*-statistic analyses. The F_{XY} figures indicate that variance ascribable to variation in the specified *X* variable (*e.g.*, population) within the specified *Y* variable (*e.g.*, spring group). Where two intermediate levels are used for a species in the hierarchy, all six cells are filled. Where one level is used only three cells are filled. The top figure in each cell is the calculated F_{XY} value and the bottom, the percentage of the total variance comprised by this.

	Variance components					
X variable:	Population	Population	Population	Level 1	Level 1	Level 2
Y variable:	Level 1	Level 2	Total	Level 2	Total	Total
Species						
Fluvidona (straight-sided)	0.226	0.434	0.426	0.269	0.257	0.015
	13	27	26	17	16	1
Fluvidona (convex) MPI 3	0.005	0.374	0.570	0.371	0.567	0.312
	0	17	26	17	26	14
MPI 4	0.152	0.539	0.679	0.456	0.622	0.305
	6	20	25	16	23	11
MPI 5	0.052	0.190	0.255	0.145	0.214	0.081
	6	20	27	15	23	9
Fonscochlea accepta	0.122		0.173		0.058	
	35		49		16	
F. aquatica	0.217	0.744	0.771	0.673	0.707	0.105
	7	23	24	21	22	3
F. zeidleri	0.346	0.677	0.718	0.506	0.568	0.127
	12	23	24	17	19	4
F. variabilis	0.382	0.599	0.781	0.352	0.646	0.454
	12	19	24	11	20	14
F. billakalina	0.101		0.220		0.133	
	22		48		29	
Trochidrobia punicea	0.118	0.262	0.574	0.164	0.518	0.423
	6	13	28	8	25	21
T. smithi	0.745		0.719		0.100	
	48		46		6	
T. minuta	0.059		0.358		0.318	
	8		49		43	
Dalhousie (globular)	0.288	0.303	0.284	0.021	−0.006	−0.027
	33	35	33	2	−1	−3
Dalhousie (pupiform)	0.304	0.363	0.384	0.085	0.115	0.032
	24	28	30	7	9	2
Dalhousie (*Fluvidona*-like)	0.538		0.656		0.049	
	43		53		4	

Nm is less than one. Values for MPI 4 are all higher than for the individual sample estimation, marginally so for the pooling of samples within tributaries and notably for the pooling into catchment based units. Even so, the data suggest that a catchment receives less than one migrant from another catchment in every three generations.

Restrictions on gene flow are also suggested by analyses of the conditional frequency of private allozymes. Estimates of migration rates based on these data are much greater than those based on F_{ST} and differ in the relative rates ascribed to the different taxa. The latter situation is particularly notable in MPI 4 which apparently has the highest rate of inter-population migration among all four species, whereas its $Nm(F_{ST})$ is the lowest.

The components of variance due to differentiation between taxonomic units at different hierarchical levels are presented in Table 2. Although comparison of these values is complicated by varying proportions of the populations being pooled at each level, some trends can be observed. Particularly striking is the concordance between the three MPI genetic species, where in each case almost half of the variation is explained by differences between tributaries or between catchments. This contrasts with the straight-sided *Fluvidona* where only one third of the variability is explained by such differences.

We have investigated gene flow in eight of the nine Lake Eyre mound springs hydrobiids, *Trochidrobia inflata* being found in only two of our sample sites. As can be seen in Figure 2 and Appendix 2, the distributions of these species vary markedly in size. *Fonscochlea accepta* is restricted to the Southern cluster of springs, *F. billakalina* to the central cluster and *T. minuta* to the northern. The range of the other species extends into more than one spring cluster, with *F. zeidleri* and *F. variabilis* being found in all three clusters. The apparent levels of gene flow between the populations of the species reflect this variability in range. *F. accepta* has high gene flow, with $Nm(F_{ST})$ between spring groups being more than two. Conversely, gene flow in *F. aquatica*, the snail which is an ecological replacement for *F. accepta* in the central

Table 3. Estimates of F_{ST} or G_{ST} in gastropods. Measures using G_{ST} (Nei, 1973) are indicated by an asterisk. Geographic scale is the distance between the extremes of the sampled range. References are: (1) Johnson and Black (1984a,b); (2) Brown (1991); (3) Mitton *et al.* (1989); (4) Campton *et al.* (1992); (5) Grant and Utter (1988); (6) Day (1990); (7) Chambers (1980); (8) Jarne and Delay (1990); (9) Mulvey *et al.* (1988); (10) Bandoni *et al.* (1990); (11) Johnson *et al.* (1988); (12) Gould and Woodruff (1986); and (13) McCracken and Brussard (1980).

Species and reference	No. of loci	No. of samples	Scale (km)	F_{ST} mean
Marine species				
(1) *Siphonaria jeanae*	4	1	10	0.002
	4	28	2,500	0.004
(2) *Haliotis rubra*	12	18	5,000	0.022
(3) *Strombus gigas*	7	21	5,000	0.076
(4) *Strombus gigas*	4	4	500	0.011*
	4	14	5,000	0.023*
(5) *Nucella lamellosa*	2	12	0.1	0.021
	2	30	1,000	0.286
(6) *Nucella lapillus*	8	6	0.5	0.015
	8	10	10	0.092
	8	15	20	0.195
Freshwater species				
(7) *Goniobasis* (2 species)	14	12	1,000	0.408
(8) *Lymnaea peregra*	6	4	50	0.018
(9) *Biomphalaria glabrata*	13	6	1,000	0.805
(10) *Biomphalaria pfeifferi*	7	12	500	0.589
Terrestrial species				
(11) *Partula taeniata*	17	22	20	0.279
P. suturalis	16	23	20	0.168
(12) *Cerion* (New Providence)	8	36	30	0.143
(13) *Triodopsis albolabris*	2	7	500	0.255

and northern clusters, is quite low, *Nm* being 25 times less between spring groups. In these large aquatic *Fonscochlea*, approximately the same relative levels of gene flow are indicated by the estimates derived from conditional allelic frequencies. Using F_{ST} for estimation, a similar pattern is shown in comparisons of the smaller aquatic species *F. billakalina* and *F. variabilis*, with gene flow between spring groups in the former being nine times the level in the latter. In *Trochidrobia*, gene flow inferred from F_{ST} statistics between spring groups in *T. minuta* is twice as high as it is in the other two species of the genus. In these latter two sets of comparisons, however, the estimates derived from conditional allelic frequencies do not show the same pattern as the F_{ST} estimates. *F. billakalina* has a similar $Nm(F_{ST})$ to *F. variabilis* and the value for *T. punicea* is almost five times as great as that for *T. minuta*.

The components of variation due to different hierarchical levels are strikingly similar in the aquatic *F. aquatica* and the amphibious *F. zeidleri*. There is some disagreement as to the level of $Nm(F_{ST})$ between spring clusters for these species, but otherwise estimates of interpopulation migration in these species are remarkably concordant. The concordance is significantly less for $Nm(\bar{p}(1))$. The components of variation are also similar in two other species (*F. variabilis* and *T. punicea*) from the Lake Eyre mound springs. Interestingly, the pattern of variation of this pair differs from that of the two large aquatic *Fonscochlea*, showing much greater between-spring divergence.

Estimated $Nm(F_{ST})$ between populations at Dalhousie Springs is in the higher reaches of the ranges observed in these studies, in both the pupiform and globular species. This trend is even more marked for inter-group migration as assayed at higher hierarchical levels. Inter-spring subgroup migration is higher in the globular snails than in any other level-one pooling, and the estimate for the pupiform snails is exceeded only by *F. accepta* and *F. billakalina*. The calculated migration rates betwen spring groups are higher for both the globular and pupiform Dalhousie radiations than for any other taxon in our studies. As expected, the proportion of variation explained by differences at the lower hierarchical levels is very high in comparison to our other studies. Indeed, virtually all of the variation in the globular snails is due to differentiation of populations within spring-subgroups, within spring groups, or within the overall spring complex (and not of spring-subgroups within groups, *etc.*). The levels of gene flow estimated from $Nm(\bar{p}(1))$ for the two higher hierarchical levels are extremely high in the context of the present results. To the extent that these estimates are credible, they reinforce the suggestion that gene flow is high in the Dalhousie Springs complex, at least as compared to the other study sites.

DISCUSSION

Our hydrobiid studies emphasise the dependence of population structure on a wide range of interacting biological and environmental factors which must be considered in historical terms. The three hydrobiid faunas we have treated extensively in this paper, and the previously studied *Tatea* (Ponder *et al.*, 1991) differ (either certainly or probably) in such biological characteristics as size, thermal and salinity tolerances and desiccation resistance. These various hydrobiids also occupy different types of habitat. *Tatea* occupies essentially continuous habitat. Wilsons Promontory *Fluvidona* and the Dalhousie Springs snails have habitats which are discontinuous, with relatively small distances between suitable areas. The Lake Eyre fauna is in discontinuous, widely separated habitat.

Before discussing our results in detail, some preliminary comparisons may be made to reinforce that gene flows actually do differ between the faunas. The Wilsons Promontory radiation occupies a much smaller geographic area than the Lake Eyre fauna, with the exception of *Fonscochlea accepta*. The markedly higher levels of gene flow in *F. accepta* might suggest that the underlying processes are more efficient in this taxon. Conversely, such relativity emphasises the very restricted levels of gene flow in the Wilsons Promontory snails. This is also suggested by comparisons of these *Fluvidona* species with the Lake Eyre *F. aquatica*, *F. variabilis* and *F. zeidleri*. The MPI 4 and MPI 3 genetic species do have a slightly higher $Nm(F_{ST})$ than the *Fonscochlea* species but this is a minor difference when the disparity between their ranges is considered. These *Fonscochlea* species each have linear extents well over an order of magnitude larger than the MPI 3 and MPI 4 *Fluvidona* (over 200km as opposed to less than 20km).

Population Structure in Quasi-Continuous Aquatic Habitats

Although the vagaries of ocean currents may reduce gene flow to or from a particular area (Todd *et al.*, 1988; Mitton *et al.*, 1989), marine snails with planktonic (and especially planktotrophic) larvae have wide natural distributions in which variation between local populations is only a minor fraction of overall genetic diversity (Table 3). This pattern is found in *Nassarius obsoletus* (Gooch *et al.*, 1972), *Littorina littorea* (Berger, 1973; Janson, 1987), *Siphonaria jeanae* (Johnson & Black, 1984a, 1984b), the species with planktotrophic larvae among the *Crepidula* studied by Hoagland (1984), *Strombus gigas* (Mitton *et al.*, 1989; Campton *et al.*, 1992) and *Haliotis rubra* (Brown, 1991) which has a lecithotrophic larva. Species distributions in these marine gastropods tend to be either widely separated, often in conjunction with geographic barriers to gene flow, or to be broadly sympatric (*e.g. Littorina* - Berger, 1973; Janson, 1987; *Crepidula* - Hoagland, 1984) presumably reflecting past allopatric speciation and subsequent dispersal into sympatry. The latter pattern was observed in our studies of *Tatea* (Ponder *et al.*, 1991). This genus, predominantly estuarine with an assumed free-swimming larval stage, has a very wide distribution, being found throughout temperate Australia. Its two species *T. rufilabris* and *T. huonensis* are sympatric over virtually all of this range. There are exceptions to these patterns of speciation in some groups with specialised feeding patterns which have high species densities (Vermeij, 1987).

Groups with direct larval development or brooding should have reduced gene flow, greater differentiation and, conceivably, higher likelihood of speciation. The first two predictions have been borne out in studies of *Littorina saxatilis* (Snyder & Gooch, 1973; Janson, 1987) and *Nucella lamellosa* (Grant & Utter, 1988). That a short planktonic phase increases rates of speciation is less certain. *Littorina saxatilis* does have a number of closely-related sibling species (Janson, 1987; Johannesson, 1988; Sundberg *et al.*, 1990) and *N. lamellosa* may represent a species complex (Grant & Utter, 1988). However, *N. lamellosa*, as presently recognised, has one of the largest geographic ranges of North Pacific gastropods.

Population Structure in Discontinuous Aquatic Habitats

There have been major investigations of gene flow and the genetic structure of two groups of the freshwater gastropods in the caenogastropod genus *Goniobasis*. Chambers (1978, 1980) investigated species from Florida and Dillon and Davis (Dillon & Davis, 1980; Dillon, 1984) those from the border regions of Virginia-North Carolina. Three main results are relevant. (1) There is a high degree of genetic divergence between populations within the same drainage system indicating low levels of gene flow. (2) There are larger differences between drainage systems, reflecting an even smaller likelihood of inter-drainage gene flow. (3) Identified taxa tend to remain allopatric or parapatric, with geographically-restricted ranges, suggesting that dispersal after speciation is limited. Dillon (1988) provides direct information on rates of gene flow in transplanted *G. proxima* populations. These are about 15–20 m upstream and 5–10 m downstream (per year), the discrepancy in movement rates being caused by the behavioural tendency of freshwater (Dillon, 1988) (and even riparian—Arter, 1990) snails to crawl upstream in compensation for down current drift.

The findings for *Goniobasis* are not true of all freshwater snails, as shown by studies of genetic variation in basommatophoran pulmonates. Dispersal in these snails is often assisted by self-fertilisation (Mimpfoundi & Greer, 1989; Bandoni *et al.*, 1990) and species distributions are generally wide-ranging. Measurement of gene flow is often hampered because of extremely low levels of genetic variation (Mimpfoundi & Greer, 1989; Jarne & Delay, 1991). Where flow can be assessed, as in *Biomphalaria straminea* (Woodruff *et al.*, 1985), *B. camerunensis* (Mimpfoundi & Greer, 1990), *B. pfeifferi* (Bandoni *et al.*, 1990) or *Bulinus cernicus* (Rollinson *et al.*,

1990), evidence of substantial population differentiation is usually found, albeit at a geographic scale much larger than in *Goniobasis*. Genetic variation in *Biomphalaria glabrata* from the Caribbean is, however, primarily due to inter-island differentiation (78%), with only 2% being due to intra-island differences (Mulvey *et al.*, 1988).

Each of the three main observations on *Goniobasis* is applicable to the Wilsons Promontory *Fluvidona* but they are less accurate descriptions of the two artesian spring faunas, indicating that they are not generally characteristic of freshwater dioecious gastropods. At Dalhousie Srings, levels of inter-spring and inter-spring group differentiation are not high and the globular and pupiform species are sympatric over a substantial range. There are instances in the Lake Eyre fauna where sister-species remain essentially allopatric (*e.g. F. accepta* and *F. aquatica*), but these are the exceptions.

Habitat Stability and Area Effects in Wilsons Promontory

Habitat on Wilsons Promontory has probably been reduced during periods of aridity. The *Fluvidona* population structure may still be showing distortions caused by recent aridity-induced interruptions of migration, particularly in the more upland MPI 3 and MPI 4 genetic species. Modelling suggests that the re-attainment of structural equilibrium following disruption is of the order of hundreds of generations rather than thousands (or more) for both the F_{ST} (Crow & Aoki, 1984) and $\bar{p}(1)$ (Slatkin & Barton, 1989) approaches. The re-attainment can, of course, only begin after the disruption is halted. In the MPI 4 genetic species there are several geographic groupings broadly definable by catchment. Gene flow between these is low, with $Nm(F_{ST})$ values of 0.148 for populations pooled within tributaries and 0.291 within catchments. Following a period of aridity, increased levels of migration in wetter times may overcome incipient genetic divergence, unless the period were so prolonged that the isolates undergo speciation. Such considerations reinforce the importance of modelling the impact of non-equilibrium population structures on measures such as F_{ST} (*e.g.* Whitlock (1992). They also suggest comparison with "area effects" and their various evolutionary consequences.

Area effects were initially recognised by their distinctive phenotypic frequency arrays sharply clinally demarcated from neighbouring areas (Cain & Currey, 1963) and were subsequently observed for allozymic frequencies (Ochman *et al.*, 1983; Johnson *et al.*, 1984). Area effects have been correlated with the rapid expansion of relict populations (Cameron & Dillon, 1984; Ochman *et al.*, 1983; Johnson *et al.*, 1984). They were initially presumed to be stable characteristics of essentially continuous populations, leading White (1978) to entertain the possibility that they might be involved in parapatric speciation. This is doubtful for *Cepaea*, at least, given that the genus contains only four species (Gould & Woodruff, 1990). Conversely, Clarke and Murray (1969) considered parapatric speciation associated with area effects was a likely cause of distribution patterns in local isolates and semispecies of the seven species of *Partula*. Although this remains a possible hypothesis (Murray & Clarke, 1980) confirmation has been hindered by the finding that, despite substantial intra-specific variability, electrophoretic divergence between reproductively-isolated taxa is low (Johnson, 1977; Johnson *et al.*, 1986a). One critical question regarding population differentiation can, however, be addressed using electrophoretic data. The studies of Johnson *et al.* (1986b) on allozymic variation suggest that, rather than deriving from multiple invasions, the fauna on Moorea evolved as an endemic radiation.

Partula is a recent invader of the Society Islands, probably arriving no more than 2.5 million years ago (Johnson *et al.*, 1986b). This contrasts with the apparent antiquity of *C. nemoralis* which can be distinguished from its sister *C. hortensis* in fossil beds at least 10 million years old (Lamotte, 1951). It is not surprising, then, that the causes of area effects, and their phylogenetic consequences, should differ between the two genera. The Wilsons Promontory *Fluvidona* exhibit a mixture of characteristics. The fauna is speciose, at least for such a small area, of probable ancient origin and subject to range expansion and contraction. In contrast to *Cepaea*, these changes have encouraged speciation. In contrast to *Partula*, this has probably been allopatric.

Speciation and Gene Flow in Lake Eyre and Dalhousie Springs

It is very difficult to predict what amount of gene flow would permit speciation because of the interdependence of biological and current and historical abiotic factors. This can be illustrated by a number of examples. Firstly, the Dalhousie Springs globular and pupiform snails are distinct species, as judged by nearly-fixed differences in wide sympatry. They are very closely related to each other and have no known close relatives. *In situ* divergence would be paradoxical if the current high levels of gene flow within these species reflect those applying historically. Comparisons of *F. zeidleri* with the allopatric pair of sister species *F. accepta* and *F. aquatica* gives a second example of this type. As judged by F_{ST} values, gene flow levels in *F. zeidleri* and *F. aquatica* are currently almost identical and very low. Why should the evolutionary fate of the ancestor of *F. accepta* and *F. aquatica* differ from that of *F. zeidleri*? Spatial or temporal local factors must have had a significant cladogenetic impact in this case.

The evolution of new species in *Partula* apparently reflects the successive west to east appearance of new land masses in the archipelago-wide basis, with no species being found on more than one island (Johnson *et al.*, 1986a). We expected a contrasting pattern in the Lake Eyre hydrobiids, owing to the potential confounding of geographic proximity with sporadic spring appearance. This was not observed, however, with almost all major genetic groups being definable by an appropriate north-

south division. Only in *F. zeidleri* and *F. billakalina* are there populations which are more closely related to distant areas than to their neighbours. This suggests that springs are colonised from relatively local sources as they arise, minimising the role of long-distance gene flow caused by factors such as bird transport.

Gene Flow Estimation by F_{ST} or $\bar{p}(1)$

Two recent studies have suggested that the private alleles model (Slatkin, 1985a) is not as useful an estimator of gene flow as F_{ST} (Waples, 1987; Johnson *et al.*, 1988). Johnson *et al.* (1988) compared the two approaches by relating them to direct estimates of gene flow in *Partula*. The private allele approach tended to give a higher estimate of interpopulation migration than did the F-statistics approach, which gave values nearer to the direct estimates. This is not always the case in our investigations; $Nm(\bar{p}(1))$ is higher than $Nm(F_{ST})$ in eight of fifteen comparisons using the overall data, eleven (of fifteen) at the first level hierarchical pooling, and eight (of ten) at the second hierarchical pooling. Waples' (1987) arguments against the general utility of the private alleles approach rest on the inconsistency of its estimates of gene flow with known larval dispersal patterns in ten species of shore fishes. In the hydrobiid data, too, there are notable inconsistencies in the estimates. For instance, $Nm(\bar{p}(1))$ for the population by population migration rate in the MPI 4 genetic species is the highest of the four Wilsons Promontory *Fluvidona*, yet in all but one other estimate, including both pooled $Nm(\bar{p}(1))$ comparisons and all $Nm(F_{ST})$, this species has the lowest estimated gene flow. The private alleles approach also gives divergent estimates of gene flow in *F. aquatica* and *F. zeidleri* while the estimates for these species from the F_{ST} approach are highly concordant.

There are several possible reasons why the private alleles model may not reflect population structure as accurately as the F_{ST} approach. Firstly, the average frequency of private alleles ($\bar{p}(1)$) may be determined from too few observations (Waples, 1987; Johnson *et al.*, 1988). The number of private alleles in our observations mainly exceeds the figure of 20, sufficient to obviate sample-size effects (Slatkin,1985a). This is so for all three calculations (population by population and both poolings) for the Wilsons Promontory MPI 4 genetic species, *F. zeidleri*, *F. aquatica*, *F. variabilis* and *T. punicea*. Sample size problems can also arise if there are too few populations and this may apply to our analyses in which data were variously pooled. Particularly, the very high estimates of $Nm(\bar{p}(1))$ in Dalhousie Springs data pooled into less than eight samples may be due to such effects although Slatkin (1985a) considered five populations sufficient to provide a good estimate.

The present data suggest that estimation problems may also be caused where there are too many samples. There are numerous instances in all three present studies of alleles which are found in only a small geographic range. Such a range, be it the length of a tributary or a small spring group, may represent the true neighbourhood size for the population. Too intensive sampling might obscure this, inflating the estimate of Nm. The impact of such effects on $\bar{p}(1)$ indicates that for this approach to be successful, there must be a fortuitous match between the scale of sampling, the biological scale of effective population boundaries and inter-population migration. Analysis of gene flow using the conditional frequency of all alleles would not be affected to anything like the same extent by the intensity of sampling. Consequently we suggest that an analytic model allowing information from all alleles to provide a numerical estimate of Nm would be more useful than current approaches.

Waples (1987) mentions some other difficulties with the private alleles approach, including the non-linearity of the estimator when $Nm > 10$ or $Nm < 0.1$ and the possibility that uniform selection may bias the estimates if $Nm < 1$ (Slatkin, 1985a). Estimation of gene flow based on F_{ST} may also be biased by the failure of the assumption that selection is negligible (Wright, 1951; Waples, 1987; Johnson *et al.*, 1988; Porter, 1990). To date, however, we have not detected significant selective differentials in our studies of hydrobiids.

Despite the apparently poorer performance of the private alleles estimates compared to the use of F-statistics, they are easy to compute and should be included in investigations of gene flow for comparative purposes. More weight should be given to F_{ST} values when grossly divergent estimates of migration rates are obtained from the two approaches. Yet such divergences should also be taken as a signal that there has been significant disturbance in the evolutionarily recent past - or that sampling does not match the scale of population structure.

ACKNOWLEDGMENTS

We thank Gerard Clark, Peter Eggler and Themo Terzis for performing the electrophoresis and collection assistance, and Janet Waterhouse, David MacIntosh, Jimmy and Hazel Ronay, Des Beechey and Roger De Keyzer for collection assistance. Alison Miller prepared the figures. We thank Professor M. Johnson, and two anonymous reviewers for attentive reading and criticism of the manuscript. This work was largely supported by grants from the Australian Research Council and the NPWS Endangered Species Program.

LITERATURE CITED

Arter, H.E. 1990. Spatial relationships and gene flow paths between populations of the alpine snail *Arianta arbustorum* (Pulmonata: Helicidae). Evolution 44:966–980.

Bandoni, S.M., M. Mulvey, D.K Koech and E.S Loker. 1990. Genetic structure of Kenyan populations of *Biomphalaria pfeifferi* (Gastropoda: Planorbidae). Journal of Molluscan Studies 56:383–391.

Barton, N.H. and M. Slatkin. 1986. A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. Heredity 56:409–415.

Berger, E. 1973. Gene-enzyme variation in three sympatric species of *Littorina*. Biological Bulletin 145:83–90.

Boeters, H.D. 1979. Species concepts of prosobranch freshwater molluscs in western Europe, 1. Malacologia 18:57–60.

Boeters, H.D. 1982. Species concepts of prosobranch freshwater molluscs in western Europe, 2. Malacologia 22:499–504.

Brown, L.D. 1991. Genetic variation and population structure in the Blacklip Abalone, *Haliotis rubra*. Australian Journal of Marine and Freshwater Reseasrch 42:77–90.

Cain, A.J. and J.D. Currey. 1963. Area effects in *Cepaea*. Philosophical Transactions of the Royal Society of London B 246: 1–181.

Cameron, R.A. and P.J. Dillon. 1984. Habitat stability, population histories and patterns of variation in *Cepaea*. Malacologia 25:271–290.

Campton, D.E., C.J. Berg, Jr., L.M. Robinson and R.A. Glazer. 1992. Genetic patchiness among populations of queen conch *Strombus gigas* in the Florida Keys and Bimini. Fishery Bulletin, U.S. 90:250–259.

Carson, H.L. and A.R. Templeton. 1984. Genetic revolutions in relation to speciation phenomena:the founding of new populations. Annual Review of Ecology and Systematics 15:97–131.

Cavalli-Sforza, L.L. and M. Feldman. 1990. Spatial subdivision of populations and estimates of genetic variation. Theoretical Population Biology 37:3–25.

Chakraborty, R. and O. Leimar. 1987. Genetic variation within a sub-divided population. *In:* Ryman, N. and F. Utter (eds.). Population genetics and fishery management. University of Washington Press, Seattle, p. 90 - 120.

Chambers, S.M. 1978. An electrophoretically detected sibling species of "*Goniobasis floridensis*" (Mesogastropoda: Pleuroceridae). Malacologia 17:157–162.

Chambers, S.M. 1980. Genetic divergence between populations of *Goniobasis* (Pleuroceridae) occupying different drainage systems. Malacologia 20:63–81.

Clarke, B. and J. Murray. 1969. Ecological genetics and speciation in land snails of the genus *Partula*. Biological Journal of the Linnean Society 1:31–42.

Cockerham, C.C. 1969. Variance of gene frequencies. Evolution 23:72–84.

Crow, J.F. and K. Aoki. 1984. Group selection for a polygenic trait: estimating the degree of population subdivision. Proceedings of the National Academy of Sciences U.S.A. 81: 6073–6077.

Davis, G. M., V. Forbes and G. Lopez. 1988. Species status of northeastern American *Hydrobia* (Gastropoda: Prosobranchia): Ecology, morphology and molecular genetics. Proceedings of the Academy of Natural Sciences of Philadelphia 140:191–246.

Davis, G. M., M. McKee and G. Lopez. 1989. The identity of *Hydrobia truncata* (Gastropoda: Hydrobiinae): Comparative anatomy, molecular genetics, ecology. Proceedings of the Academy of Natural Sciences of Philadelphia 141: 333–359.

Day, A.J. 1990. Microgeographic variation in allozyme frequencies in relation to the degree of exposure to wave action in the dogwhelk *Nucella lapillus* (L.) (Prosobranchia: Muricacea) Biological Journal of the Linnean Society 40:245–261.

Dillon, R.T., Jr. 1984. Geographic distance, environmental difference, and divergence between isolated populations. Systematic Zoology 33:69–82.

Dillon, R.T., Jr. 1988. The influence of minor human disturbance on biochemical variation in a population of freshwater snails. Biological Conservation 43:137–144.

Dillon, R.T., Jr. and G.M. Davis. 1980. The *Goniobasis* of Southern Virginia and Northwestern North Carolina: Genetic and shell morphometric relationships. Malacologia 20:83–98.

Endler, J.A. 1973. Gene flow and population differentiation. Science 179:243–250.

Felsenstein, J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5:164–166.

Gooch, J.L., B.S. Smith and D. Knupp. 1972. Regional survey of gene frequencies in the mud snail *Nassarius obsoletus*. Biological Bulletin 142:36–48.

Gould, S.J. and D.S. Woodruff. 1986. Evolution and systematics of *Cerion* (Mollusca: Pulmonata) on New Providence Island: a radical revision. Bulletin of the American Museum of Natural History 182:389–490.

Gould, S.J. and D.S. Woodruff. 1990. History as a cause of area effects: an illustration from *Cerion* on Great Inagua, Bahamas. Biological Journal of the Linnean Society 40: 67–98.

Govindaraju, D.R. 1989. Estimates of gene flow in forest trees. Biological Journal of the Linnean Society 37:345–357.

Grant, V. 1980. Gene flow and the homogeneity of species populations. Biological Zhentralblatt 99:157–169.

Grant, W.S. and F.M. Utter. 1988. Genetic heterogeneity on different geographic scales in *Nucella lamellosa* (Prosobranchia, Thaididae). Malacologia 28:275–287.

Habermehl, M.A. 1982. Springs in the Great Artesian Basin, Australia - their nature and significance. Bureau of Mineral Resources, Canberra, Geology and Geophysics, Report No. 235.

Harris, C. 1993. Mound springs: South Australian conservation inititives. Rangelands Journal 14:157–173

Hebert, P.D.N. and M.J. Beaton. 1989. Methodologies for Allozyme analysis using Cellulose Acetate Electrophoresis. Helena Laboratories, Austin. 32 pp.

Hoagland, K.E. 1984. Use of molecular genetics to distinguish species of the gastropod genus *Crepidula* (Prosobranchia: Calyptreidae). Malacologia 25:607–628.

Janson, K. 1987. Allozyme and shell variation in two marine snails (*Littorina*, Prosobranchia) with different dispersal abilities. Biological Journal of the Linnean Society 30:245–256.

Jarne, P. and B. Delay. 1990. Population genetics of *Lymnaea peregra* (Müller) (Gastropoda: Pulmonata) in Lake Geneva. Journal of Molluscan Studies. 56:317–321.

Jarne, P. and B. Delay. 1991. Population genetics of freshwater snails. Trends in Ecology and Evolution 6:383–386.

Johanessen, K. 1988. The paradox of Rockall: Why is a brooding gastropod (*Littorina saxatilis*) more widespread than one having a planktonic larval dispersal stage (*L. littorea*)? Marine Biology 99:507–513.

Johnson, M.S. and R. Black. 1984a. The Wahlund effect and the geographic scale of variation in the intertidal limpet *Siphonaria* sp. Marine Biology 70:295–302.

Johnson, M.S. and R. Black. 1984b. Pattern beneath the chaos: The effect of recruitment on genetic patchiness in an intertidal limpet. Evolution 38:1371–1383.

Johnson, M.S., B. Clarke and J. Murray. 1977. Genetic variation and reproductive isolation in *Partula*. Evolution 31: 116–126.

Johnson, M.S., J. Murray and B. Clarke. 1986a. An electrophoretic analysis of phylogeny and evolutionary rates in

the genus *Partula* from the Society Islands. Proceedings of the Royal Society of London B 227:161–177.

Johnson, M.S., J. Murray and B. Clarke. 1986b. High genetic similarities and low hetreozygosities in land snails of the genus *Samoana* from the Society Islands. Malacologia 27: 97–106.

Johnson, M.S., B. Clarke and J. Murray. 1988. Discrepancies in the estimation of gene flow in *Partula*. Genetics 120: 233–238.

Johnson, M.S., O.C. Stine and J. Murray. 1984. Reproductive compatibility despite large-scale genetic divergence in *Cepaea nemoralis*. Heredity 53:655–665.

Kabat, A.R. and R. Hershler. 1993. The prosobranch snail family Hydrobiidae (Gastropoda: Rissooidea): Review of classification and supraspecific taxa. Smithsonian Contributions to Zoology. Smithsonian Contributions to Zoology, 547:1–94.

Kimura, M. and G.H. Wiess. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561–576.

Kotwicki, V. 1989. Floods in the western Lake Eyre basin and their impact on Dalhousie Springs. *In:* Zeidler, W. and W.F. Ponder (eds.). Natural history of Dalhousie Springs. South Australian Museum, Adelaide, p. 41–45.

Krieg, G.W. 1989. Geology. *In:* Zeidler, W. and W.F. Ponder (eds.). Natural history of Dalhousie Springs. South Australian Museum, Adelaide, p. 19–26.

Lamotte, M. 1951. Recherches sur la structure génétique des populations naturelles de *Cepaea nemoralis* L. Bulletin Biologique de France, Supplement 35:1–239.

Lassen, H.H. 1979. Electrophoretic enzyme patterns and breeding experiments in Danish mud snails (Hydrobiidae). Ophelia 318:38–87.

McCracken, G.F. and P.F. Brussard. 1980. The population biology of the White-lipped land snail, *Triodopsis albolabris*: genetic variability. Evolution 34:92–104.

Mimpfoundi, R. and G.J. Greer. 1989. Allozyme comparisons among species of the *Bulinus forskalii* group (Gastropoda: Planorbidae) in Cameroon. Journal of Molluscan Studies 55:405–410.

Mimpfoundi, R. and G.J. Greer. 1990. Allozyme variation among populations of *Biomphalaria camerunensis* (Boettger, 1941) (Gastropoda: Planorbidae) in Cameroon. Journal of Molluscan Studies 56:373–381.

Mitton, J.B., C.J. Berg, Jr., and K.S. Orr. 1989. Population structure, larval dispersal and gene flow in the Queen Conch, *Strombus gigas* of the Caribbean. Biological Bulletin 177:356–362.

Mulvey, M., M.C. Newman and D.S. Woodruff. 1988. Genetic differentiation among West Indian populations of the schistosome-transmitting snail *Biomphalaria glabrata*. Malacologia 29:309–317.

Murray, J. and B. Clarke. 1980. The genus *Partula* on Moorea: speciation in progress. Proceedings of the Royal Society of London B 211:83–117.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences U.S.A. 70:3321–3323.

Nei, M. and R.K. Chesser. 1983. Estimation of fixation indices and gene diversification. Annals of Human Genetics 47: 253–259.

Ochman, H., J.S. Jones and R.K. Selander. 1983. Molecular area effects in *Cepaea*. Proceedings of the National Academy of Sciences U.S.A. 80:4189–4193.

Ponder, W.F. 1986. Mound springs of the Great Artesian Basin. *In:* DeDeckker, P. and W.D. Williams (eds.). Limnology in Australia CSIRO, Melbourne and W. Junk, The Hague, p. 403–420.

Ponder, W.F. 1989. Mollusca. *In:* Zeidler, W. and W.F. Ponder (eds.). Natural history of Dalhousie Springs. South Australian Museum, Adelaide, p. 71–80.

Ponder, W. F. and G.A. Clark. 1988. A morphological and electrophoretic examination of '*Hydrobia buccinoides*', a variable brackish-water gastropod from temperate Australia (Mollusca: Hydrobiidae). Australian Journal of Zoology 36:661–689.

Ponder, W.F., D.J. Colgan and G.A. Clark. 1991. The morphology, taxonomy and genetic structure of *Tatea* (Mollusca: Gastropoda: Hydrobiidae), estuarine snails from temperate Australia. Australian Journal of Zoology 39:447–497.

Ponder, W. F., R. Hershler and B. Jenkins. 1989. An endemic radiation of hydrobiid snails from artesian springs in northern South Australia: their taxonomy, physiology, distribution and anatomy. Malacologia 31:1–140.

Ponder W.F., D.J. Colgan, G.A. Clark, A.C. Miller and T. Terzis. 1994. Microgeographic genetic and morphological differentiation of freshwate snails - a study on the Hydrobiidae of Wilsons Promontory, Victoria, southeastern Australia. Australian Journal of Zoology, In press.

Porter, A.H. 1990. Testing nominal species boundaries using gene flow statistics: the taxonomy of two hybridising Admiral butterflies (*Limenitis*: Nymphalidae). Systematic Zoology 39:131–147.

Preziosi, R.F. and D.J. Fairbairn. 1992. Genetic population structure and levels of gene flow in the stream-dwelling waterstrider, *Aquarius* (= *Gerris*) *remigis* (Hemiptera: Gerridae). Evolution 46:430–444.

Rees, W.J. 1965. The aerial dispersal of Mollusca. Proceedings of the Malacological Society of London 36:269–282.

Rohlf, F. J. 1990. NTSYS-pc. Numerical taxonomy and multivariate analysis system. Version 1.60. Exeter Software, New York.

Rollinson, D., R.A. Kane, A. Warlow and V.R. Southgate. 1990. Observations on genetic diversity of *Bulinus cernicus* (Gastropoda: Planorbidae) from Mauritius. Journal of Zoology 222:19–26.

Schmidt, E.R. and I.W.B. Thornton. 1992. The Psocoptera (Insecta) of Wilsons Promontory National Park, Victoria. Memoirs of the Museum of Victoria, 53, 137–220.

Skibinski, D.O.F., J.A. Beardmore and T.F. Cross. 1983. Aspects of the population genetics of *Mytilus* (Mytilidae: Mollusca) in the British Isles. Biological Journal of the Linnean Society 19:137–183.

Slatkin, M. 1981. Estimating levels of gene flow in natural populations. Genetics 99:323–335.

Slatkin, M. 1985a. Rare alleles as indicators of gene flow. Evolution 39:53–65.

Slatkin, M. 1985b. Gene flow in natural populations. Annual Review of Ecology and Systematics 16:393–430.

Slatkin, M. and N.H. Barton. 1989. A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43:1349–1368.

Smith, P.C. 1989. Hydrogeology. *In:* Zeidler, W. and W.F. Ponder (eds.). Natural history of Dalhousie Springs. South Australian Museum, Adelaide, p. 27–39.

Snyder, T.P. and J.L. Gooch. 1973. Genetic differentiation in *Littorina saxatilis* (Gastropoda). Marine Biology 22:177–182.

Sundberg, P., A.J. Knight, R.D. Ward and K. Johannesson.

1990. Estimating the phylogeny in mollusc *Littorina saxatilis* (Olivi) from enzyme data: methodological considerations. Hydrobiologia 193:29–40.

Swofford, D.L. and R.B. Selander. 1981. BIOSYS–1: a FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. Journal of Heredity. 72: 281–283.

Thomson, R. and S. Barnett. 1985. Geology, geomorphology and hydrogeology. *In:* Greenslade, J., L. Joseph and A. Reeves (eds.). South Australia's mound springs. Nature Conservation Society of South Australia, Adelaide, p. 3–26.

Todd, C.D., J.N. Havenhand and J.P. Thorpe. 1988. Genetic differentiation, pelagic larval transport and gene flow between local populations of the intertidal marine mollusc *Adalaria proxima* (Alder and Hancock). Functional Ecology 2:441–451.

Vermeij, G. 1987. The dispersal barrier in the tropical Pacific: implications for molluscan speciation and extinction. Evolution 41:1046–1058.

Wallis, G.L. 1988. Wilsons Promontory. *In:* Clark, I. and B. Cook (eds.). Victorian Geology Excursion Guide. Australian Academy of Science, Canberra, 353–372.

Waples, R.S. 1987. A multispecies approach to the analysis of gene flow in marine shore fishes. Evolution 41:385–400.

Weir, B. 1990. Intraspecific differentiation. *In:* Hillis, D.M. and C. Moritz (eds.). Molecular systematics. Sinauer Associates, Sunderland, p. 373–410.

Weir, B. and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

White, M.J.D. 1978. Modes of speciation. W. H. Freeman and Co., San Francisco. pp. 455.

Whitlock, M.C. 1992. Temporal fluctuations in demographic parameters and the genetic variance among populations. Evolution 46:608–615.

Williams, A.F. and J.W. Holmes. 1978. A novel method of estimating the discharge of water from mound springs of the Great Artesian Basin, Central Australia. Journal of Limnology 38:262–272.

Woodruff, D.S., M. Mulvey and M.W. Yipp. 1985. Population genetics of *Biomphalaria straminea* in Hong Kong. Journal of Heredity 76:355–360.

Wool, D. 1987. Differentiation of island populations: a laboratory model. American Naturalist 129:188–202.

Wopfner, H. and C.R. Twidale. 1967. Geomorphological history of the Lake Eyre Basin. *In:* Jennings, J.N. and J.A. Mabbutt (eds.). Landform studies in Australia . Australian National University Press, Canberra, p. 118–143.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics 15:323–354.

Wright, S. 1978. Evolution and the genetics of populations. Vol. 2. The theory of gene frequencies. University of Chicago Press, Chicago.

Zeidler, W. and W.F. Ponder. 1989. Natural history of Dalhousie Springs. South Australian Museum, Adelaide. pp. 138.

Appendix 1. The numbers of interpretable loci in the taxonomic groups of studied snails. The columns give the enzyme name, its abbreviation, E.C. number and the number of loci which were interpretable in the designated taxa. Column A is Wilsons Promontory *Fluvidona*; B is *Fonscochlea accepta*; C *F. aquatica*; D *F. zeidleri*; E *F. variabilis*; F *Trochidrobia punicea*; G *T. smithi* and *T. minuta*; H, the Dalhousie Springs globular and pupiform snails; and I, the Dalhousie *Fluvidona*-like species.

Enzyme	Abbreviation	E.C. no.	Loci								
			A	B	C	D	E	F	G	H	I
Alkaline phosphatase	ALKP	3.1.3.1								1	1
Aspartate aminotransferase	AAT	2.6.1.1	2	2	2	2	2	2	1	2	2
Enolase	ENO	4.2.1.11	1	1	1	1	1	1	1	1	1
Esterase	EST	3.1.1.1	1	3	2	3	3	3	1	3	3
Fumarate hydratase	FH	4.2.1.2		1	1	1	1	2		1	
β-Galactosidase	GAL	3.2.1.23								1	
Glycerol-3-phosphate dehydrog.	GPD	1.1.1.8					1	1	1	1	1
Glucosephosphate isomerase	GPI	5.3.1.9	1	1	1	1	1	1	1	1	1
Glutamate-pyruvate transaminase	GPT	2.6.1.2	2	1	2	2	2		2		2
Hexokinase	HK	2.7.1.1		1	1	1					
Hydroxybutyrate dehydrogenase	HBDH	1.1.1.30						1	1	1	1
Isocitrate dehydrogenase	IDH	1.1.1.42	2	2	2	1	2	2	2	2	2
Leucine aminopeptidase	LAP	3.4.11	1							1	
Malate dehydrogenase	MDH	1.1.1.37	2	2	1	2	2	2	2	2	2
Mannosephosphate isomerase	MPI	5.3.1.8	1	1	1	1	2	1	1	2	1
Octopine dehydrogenase	ODH	1.5.1.15		1	1	1	1				
Peptidases: Phe-Pro substrate	PPR	3.4.11	1							1	1
: Leu-Ala substrate	LAL	3.4.11	2					3	2	1	
: Val-Leu substrate	VL	3.4.11		1	1	1	2				
: Leu-Leu-Gly substrate	LGG	3.4.11		1	1	1	1			2	
Phosphoglucomutase	PGM	5.4.2.2	2	2	2	1	1	1	1	2	2
6-Phosphogluconate dehydrog.	6-PGDH	1.1.1.44	1	1	1	1	1	1	1	1	1
Purine nucleoside phosphorylase	PNP	2.4.2.1		1		1	1				
Sorbitol dehydrogenase	SDH	1.1.1.14	2	1	2	1	2	2			2
Triosephosphate isomerase	TPI	5.3.1.1	1	1	1	1	1	1	1	1	1
UDP glucose pyrophosphorylase	UDPG	2.7.7.9								1	1

Appendix 2. Hierarchical sample structure analysed for various hydrobiid species. The highest hierarchical levels are written flush to the left margin of each column. Lower levels are successively indented. Spring groups and watersheds are identified in figures 1 to 3. The average number of specimens scored for each locus is given after the sample designations.

Wilsons Promontory *Fluvidona* Hierarchy: Catchment-Stream-Site. 'MAIN' indicates the principal stream of the catchment, 'ONE', 'TWO', *etc.*, the tributaries. Sites within catchments are numbered, approximately, clockwise.

Straight-sided	
WHISKY CREEK	
LOWER	
WC1	12.64
WC3	8.36
UPPER	
WC5	12.09
WC6	8.27
WC7	11.41
SQUEAKY CREEK	
MAIN	
SC2	7.09
FRASER CREEK	
MAIN	
FC1	7.09
GROWLER CREEK	
MAIN	
GC10	7.55
Convex MP13	
GROWLER CREEK	
MAIN	
GC12	7.27
FIVE	
GC13	7.00
FRASER CREEK	
MAIN	
FC3	7.54
ROARING MEG	
MAIN	
RM1	7.54
RM3	8.54
ONE	
RM2	5.27
TWO	
RM4	13.64
FIRST BRIDGE CREEK	
MAIN	
FB1	7.54
PICNIC CREEK	
MAIN	
PC1	7.32

Appendix 2. Continued.

FRESHWATER CREEK	
TWO	
FW3	6.41
WATERLOO BAY	
MAIN	
WB1	7.09
WB2	7.09
Convex MP14	
WHISKY CREEK	
MAIN	
WC6	16.09
WC7	13.55
WC8	12.95
WC9	12.99
WC10	13.95
WC11	13.90
WC12	27.95
TIDAL RIVER	
ONE	
TR4	38.45
TR5	8.63
THREE	
TR7	8.36
TR8	7.63
FOUR	
TR9	8.00
FIVE	
TR11	8.09
TR10	8.81
TWO	
TR2	8.63
SIX	
TR6	8.36
GROWLER CREEK	
MAIN	
GC21	11.41
ONE	
GC16	8.96
TWO	
GC6	7.93
THREE	
GC4	7.82
GC5	7.91
BLACKFISH CREEK	
ONE	
BC4	5.91
TWO	
BC6	5.91
THREE	
BC8	7.45

Appendix 2. Continued.

SQUEAKY CREEK	
MAIN	
SC2	7.36
ONE	
SC3	7.36
Convex MPI5	
CHINAMANS CREEK	
MAIN	
CC1	15.72
DARBY RIVER	
MAIN	
DR2	14.50
ONE	
DR3	13.96
TWO	
DR4	14.64
WHISKY CREEK	
MAIN	
WC1	12.55
WC2	10.00
WC4	11.64
WC5	11.64
TIDAL RIVER	
MAIN	
TR1	10.72
GROWLER CREEK	
MAIN	
GC25	3.34
FRESHWATER CREEK	
MAIN	
FW1	7.09
FRASER CREEK	
MAIN	
MN1	7.09
ONE	
FC1	7.09
SQUEAKY CREEK	
MAIN	
SC1	14.28

Lake Eyre Hydrobiidae Hierarchy: Region-Spring group-Site. Samples are identified by spring group initials and a number indicating south–north order in the entire Lake Eyre collections.

F. accepta	
WELCOME SPRINGS	
WS1	15.38
WS2	10.38
WS3	15.38

Appendix 2. Continued.

DAVENPORT SPRINGS	
DS4	13.92
DS5	10.38
DS6	13.62
HERMIT HILL	
HH7	15.42
HH8	10.38
HH9	15.38
HH10	10.46
HH11	10.46
HH12	10.42
F. aquatica	
MIDDLE	
BLANCHE CUP	
BC13	9.70
BC15	9.85
BC16	9.85
COWARD/KEWSON	
CS19	9.81
ES20	9.15
ES21	9.77
KH27	7.54
JS28	9.73
JE29	9.31
BERESFORD SPRINGS	
BS22	9.69
STRANGWAYS SPRINGS	
SS24	8.85
SS30	6.77
NORTH	
OUTSIDE SPRINGS	
OS25	8.85
TM26	9.00
FREELING SPRINGS	
FR31	10.73
FR32	9.73
F. billakalina	
BILLK	
BK18	7.92
STRANGWAYS	
SS24	8.00
SS30	9.13
F. variabilis	
SOUTH	
WELCOME SPRINGS	
WS2	9.80
DAVENPORT SPRINGS	
DS6	8.60
MIDDLE	
BLANCHE CUP	
BC13	9.70
BC14	10.44
BC15	10.20
BC16	7.92

Appendix 2. Continued.

COWARD/KEWSON	
CS17	7.92
CS19	8.00
ES20	8.56
ES21	10.08
KH27	8.44
JS28	10.20
JE29	7.80
BERESFORD SPRINGS	
BS22	8.32
WA23	8.20
NORTH	
OUTSIDE SPRINGS	
OS25	8.52
TM26	8.00
FREELING SPRINGS	
FR31	7.75
FR32	7.96
F. zeidleri	
SOUTH	
HERMIT SPRINGS	
HH8	8.50
MIDDLE	
BLANCHE CUP	
BC13	7.54
BC14	10.25
COWARD/KEWSON	
CS17	10.25
CS19	8.75
ES20	10.21
ES21	9.96
KH27	7.88
JS28	10.17
JE29	8.33
BILLAKALINA	
BK18	8.75
BERESFORD SPRINGS	
BS22	10.04
WA23	10.25
STRANGWAYS SPRINGS	
SS24	10.09
NORTH	
OUTSIDE SPRINGS	
TM26	8.88
FREELING SPRINGS	
FR32	8.79
T. punicea	
SOUTH	
WELCOME SPRINGS	
WS1	8.00
WS3	9.46

Appendix 2. Continued.

DAVENPORT SPRINGS	
DS4	9.00
DS5	8.92
DS6	8.00
HERMIT HILL	
HH7	8.00
HH8	8.00
HH10	7.96
HH12	8.00
MIDDLE	
BLANCHE CUP	
BC13	8.00
BC15	8.08
COWARD/KEWSON	
CS19	7.38
ES20	7.88
ES21	7.92
KH27	7.88
JS28	6.92
JE29	7.92
T. smithi	
MIDDLE	
BLANCHE CUP	
BC14	14.00
BILLAKALINA	
BK18	16.89
BERESFORD SPRINGS	
BS22	10.22
STRANGWAYS SPRINGS	
SS24	13.27
SS30	9.22
NORTH	
OUTSIDE SPRINGS	
OS25	15.33
TM26	11.89
T. minuta	
OUTSIDE SPRINGS	
OS25	8.44
FREELING SPRINGS	
FR31	13.61
FR32	11.56

Dalhousie Springs Hydrobiidae Hierarchy: Spring group-Subgroup-Site. Samples marked with a "p" are from the pool of large springs and those with "o" from the outflow.

Globular			
A	ONE	A1	6.64
		A3	8.57
		A8	19.57
B	ONE	B1	11.43
C	A	Calp	49.50
		Calo	37.07
		Calo	28.39
		Calo	33.54
		Ca9	32.25

Appendix 2. Continued.

	B	Cb2	6.93
		Cb2a	5.93
		Cb2b	6.29
	C	Cc1	6.07
		Cc3	6.50
	D	Cdlp	7.40
		Cdlp	7.00
		Cdlo	4.32
		Cdlo	7.00
		Cd2	14.25
		Cd8	5.21
Pupiform			
A	ONE	A1	18.89
		A2	21.29
		A3	6.11
		A6	17.82
		A8	34.11
B	ONE	B1	8.86
		B2	9.79
C	A	Ca2	29.79
		Ca3	26.29
		Ca5	33.25
		Ca7a	28.35
		Ca7b	30.71
		Ca8	17.79
		Ca12	30.18
		Ca13	28.25
	B	Cb4	8.21
		Cb5o	6.00
		Cb5p	17.86
	C	Cc1	6.43
		Cc4	16.43
		Cc8	8.32
	D	Cdlp	6.96
		Cd3	5.61
		Cd5	5.43
		Cd9	6.36

Appendix 2. Continued.

D	A	Da1	6.00
		Da2	17.64
		Da3	11.75
	B	Db1	6.00
		Db2	9.79
		Db4	5.96
E	A	E5	23.21
	B	E1	19.85
		E2	9.21
		E7a	18.07
		E8	21.93
F	ONE	F1	20.04
		F2	17.89
G	ONE	Ga2	26.07
		Ga3	33.14
		Ga4	15.00
		Ga6a	6.39
		Ga6b	14.07
H	ONE	H1	6.96
	TWO	H3	17.75
Fluvidona-like			
C		Cd11	4.16
		Elo	8.32
E		E3	4.16
F		F9	4.16
G		Ga1	4.16
		Ga2o	4.16
		Ga6	4.16

Allozyme Cladistics in Malacology: Why and How?

Kenneth C. Emberton
Department of Malacology
Academy of Natural Sciences of
Philadelphia
1900 Benjamin Franklin Parkway
Philadelphia, PA 19103-1195, USA

ABSTRACT

A hypothetical but plausible data set is introduced for which Mickevich and Mitter's (1981, 1983) qualitative-coding, minimum-turnover cladistic (= discrete parsimony) method accurately reconstructs phylogeny, whether rare alleles are detected or not, but for which both the UPGMA distance method and a hand-calculated application of Swofford and Berlocher's (1987) frequency-parsimony method give incorrect phylogenies. Mickevich and Mitter's (1981, 1983) method is outlined and demonstrated. When applying this method to 20 polygyrid genera using Hennig86 (Farris, 1988), two problems arose and were circumvented. First, allelic combinations occurred in complexly interrelated sets (interim solution: treat such sets as single character-states); and second, alternative character-state trees existed for each locus (solution: binary-code each alternative, then weight by the reciprocal of the number of alternatives). Cladistic analysis of the polygyrid-genera allozyme data (Emberton, 1994) ordered yielded the same topology as, but higher resolution than, when run unordered.

Key words: allozymes, phylogenetics, cladistics, discrete parsimony, distance methods, frequency-parsimony, ordered vs. unordered data, Gastropoda Polygyridae.

INTRODUCTION

Allozyme data remain among the easiest and cheapest to obtain for molecular systematics (Richardson *et al.*, 1986; Hillis & Moritz, 1990). Despite the current revolution in nucleic-acid sequencing (see other papers in this volume), allozymes may continue to play a major role in molluscan systematics in many laboratories throughout the world, for many years to come.

Of the two discrete ways to use allozyme data for systematics (Sarich, 1977), population genetics has dominated in molluscan studies (Berger, 1983; Cain, 1983; Johnson *et al.*, 1988; Hillis, 1989; Woodruff, 1989; Woodruff & Solem, 1990) while phylogenetic reconstruction has been less common. This paper deals only with phylogenetic reconstruction.

There are three main approaches to phylogenetic reconstruction using allozymes (Buth, 1984; Swofford & Berlocher, 1987): distance methods (Sneath & Sokal, 1973; Farris, 1972), cladistics (= discrete parsimony = phylogenetic but not taxonomic aspects of cladistics) (Hennig, 1966; Wiley, 1981; Brooks & McLennan, 1991; Harvey & Pagel, 1991), and frequency-parsimony (Swofford & Berlocher, 1987). Of these, distance methods have been by far the most commonly used in molluscan systematics (e.g. Davis *et al.*, 1981; Johnson *et al.*, 1986; Emberton, 1988), whereas cladistics has been relatively uncommon (Emberton, 1988, 1991; Hoeh, 1990) and frequency-parsimony remains untried. Which of these three is the most appropriate application of allozyme data to phylogenetic reconstruction? What is the best method? Does this method hold up even when rare alleles are undetected? Is this method practical? What are its pitfalls, and how can they be avoided?

The purpose of this paper is to begin to address these questions by (1) introducing a hypothetical case of allozyme evolution for which only a cladistic method, and neither UPGMA (a distance method) nor frequency-parsimony (as understood and hand-calculated by the present author), accurately reconstructs phylogeny, both with and without detection of rare alleles; (2) explaining and demonstrating Mickevich & Mitter's (1981, 1983) qualitative-coding, minimum-turnover cladistic method; and (3) documenting how this method was applied, using Hennig86 programs (Farris, 1988), to a complex malacological data set (Emberton, 1994).

MATERIALS AND METHODS

The hypothetical phylogeny (Figure 1) consists of an outgroup (out) and three taxa (A, B, and C) with the tree topology: out(A(B,C)). The devised allozyme data from this phylogeny (Figure 1) consist of three loci (locus 1, locus 2 and locus 3), each of which has three alleles (a, b, and c). In all three loci the designated course of evolution was from allele a to allele b, passing through the intermediate stage of heterozygosity ab.

In locus 1, the outgroup is fixed for a, taxon A is heterozygous for ab, and taxon B is fixed for b; taxon C has evolved further to acquire a third allele (c), for which it is heterozygous (bc). Locus 2 in this hypothetical case

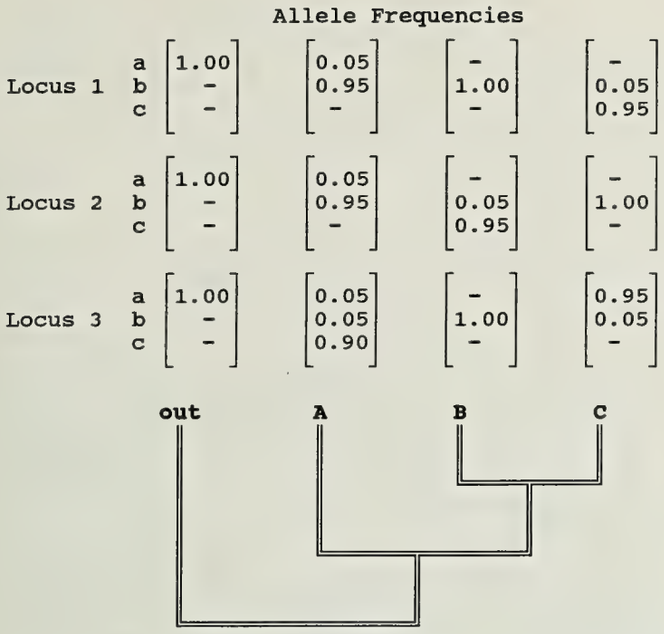


Figure 1. Hypothetical phylogeny and allozyme frequencies for three loci frequencies for three loci. See text for explanation.

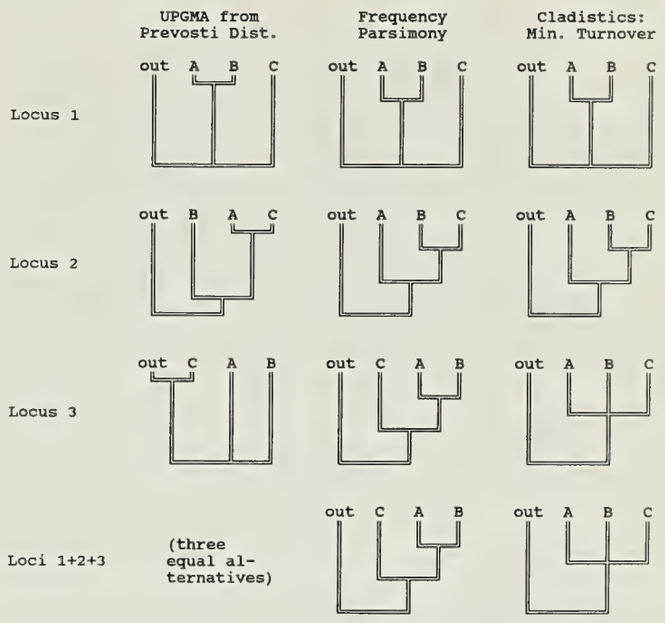


Figure 3. Same as Figure 2, but after deleting allelic occurrences at frequencies of 0.05.

is identical to locus 1, except that allele c is acquired by taxon B rather than by taxon C. Locus 3 differs in that allele c is acquired only by taxon A, and taxon C partially reverts to allele a. The frequencies of alleles are shown in Figure 1, and range from 0.05 to 1.00.

For a representative distance-method analysis of this hypothetical phylogeny and allozyme-evolution pattern, unweighted pair group, mathematical averaging (UPGMA) was used (Sneath & Sokal, 1973). This method can be applied to any genetic-distance matrix; the index of genetic distance chosen here was that of Prevosti (see Wright, 1978) because of its simplicity of calculation, its

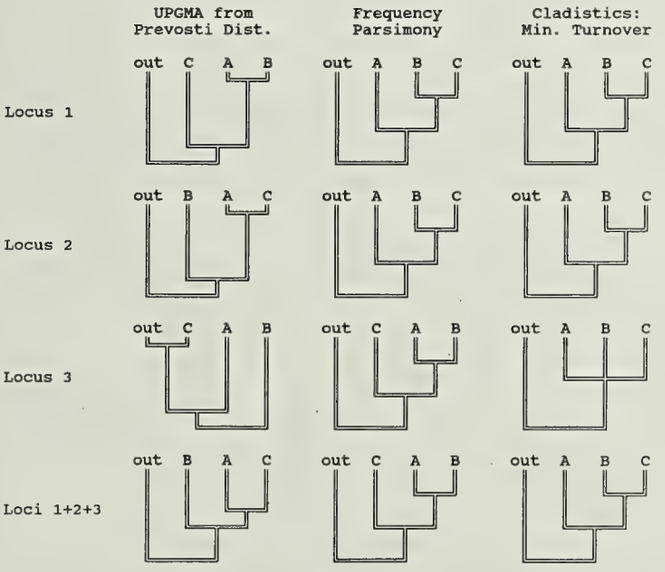


Figure 2. Performance tests of three different methods of phylogenetic inference, analysing the hypothetical data set in Figure 1. See text for details.

meeting of the triangle-equality criterion, and its similarity in performance to the preferred—both theoretically and empirically—Cavalli-Sforza and Edwards arc and arc-chord indices (see Wright, 1978). It must be emphasized that UPGMA, although commonly used, is well known to be highly prone to inaccuracies when evolutionary rates differ among lineages. A fuller and fairer test of distance methods—beyond the scope of this paper—would have to include neighbor joining and other methods that do not assume constancy of evolution in all lineages.

Application of frequency-parsimony attempted to follow Swofford and Berlocher (1987), analyzing the data by hand rather than using the FREQPARS program of Swofford (1988). According to the present author's understanding, the frequency-parsimony method (Swofford & Berlocher, 1987) finds a set of hypothetical ancestors that minimizes the total amount of change in all alleleic frequencies; there may be more than one set of hypothetical ancestors—in such cases, hypothetical ancestors were chosen as identical to extant taxa whenever possible.

Cladistic (= discrete-parsimony) analysis treated the locus as character, allelic combinations as character-states, and heterozygotes as evolutionarily intermediate between homozygotes for the same alleles (qualitative coding, minimum turnover model of Mickevich & Mitter, 1981, 1983; see below). This is a favored method of cladistic treatment of allozyme data (Buth, 1984), but a fuller evaluation would also have to consider transmodal theories as outlined by Mickevich and Weller (1990).

Each of these three methods of phylogeny reconstruction was applied four times: to locus 1, to locus 2, to locus 3, and to all three loci combined. The entire analysis was then repeated under the assumption that rare alleles were

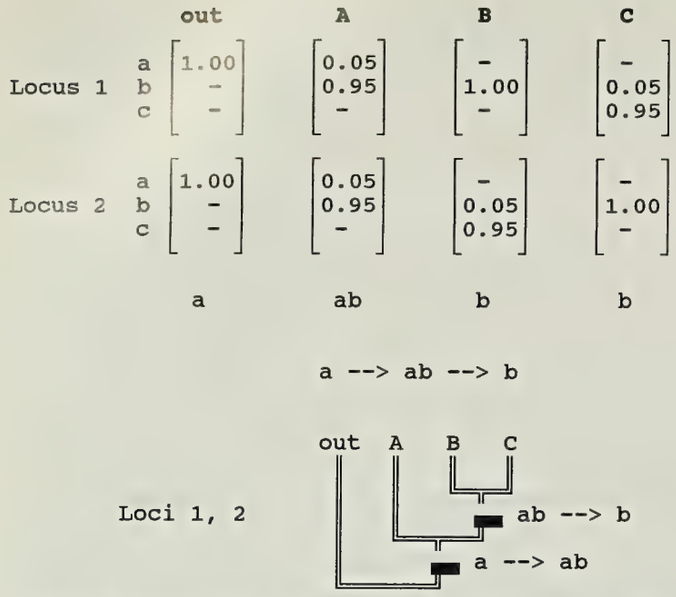


Figure 4. Application of cladistic steps 1–5 (see text) to Figure 1's loci 1 and 2. See text for explanation.

undetected, i.e. after deleting all allelic occurrences of frequency 0.05. All computations were by hand.

To explain and demonstrate Mickevich and Mitter's (1981, 1983) qualitative-coding, minimum-turnover cladistic method, the method was broken down into easy-to-follow steps, with a worked-through example using the hypothetical data set (Figure 1).

This method was applied, using Hennig86 programs (Farris, 1988), to a data set from 20 genera of polygyrid land snails (Gastropoda: Pulmonata: Stylommatophora), as part of a broad-based phylogenetic analysis that also incorporated behavior and reproductive anatomy (Emberton, 1994). The data consisted of eight cladistically informative loci whose allelic variation was classified into 29 character states. Hennig86 programs were used because Platnick's (1989) empirical tests found them superior to any other programs then available in finding the most parsimonious cladograms from real data sets. Current PAUP programs, however, may perform more accurately than in the past, and they would probably be more flexible in dealing with the polygyrid data set (see below).

RESULTS

Figure 2 shows results of the three methods applied to the data of Figure 1. UPGMA gave the incorrect phylogeny for each of the three loci, as well as for combined loci. Frequency-parsimony (as interpreted by the present author) gave the correct phylogeny for loci 1 and 2, but the incorrect phylogeny both for locus 3 and for all three loci combined. The cladistic method, on the other hand, always gave the correct phylogeny, although with incomplete resolution for locus 3, in which taxa A, B, and C appear in a trichotomy.

Deleting rare alleles gave similar results, but with less resolution overall (Figure 3). Thus UPGMA produced

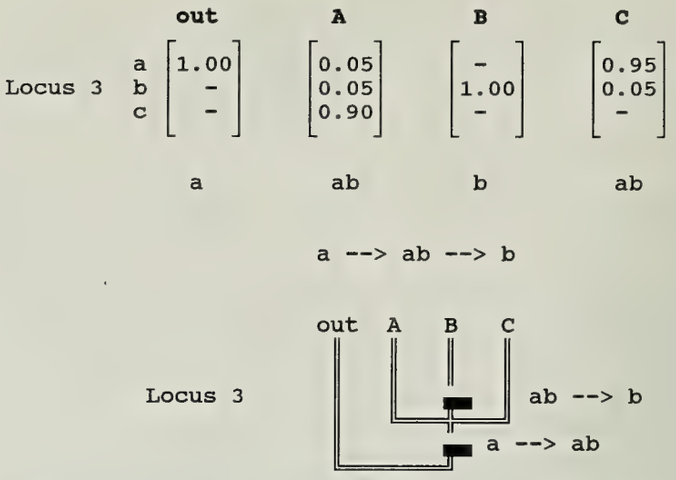


Figure 5. Application of cladistic steps 1–5 (see text) to Figure 1's locus 3. See text for explanation.

the incorrect phylogeny for all three loci, and gave three equal alternative phylogenies for the combined loci. Frequency parsimony (as interpreted by the present author) was correct for locus 2, but gave incorrect results for loci 1 and 3 and for combined loci. The cladistic method gave topologies that were incorrect for locus 1, correct and resolved for locust 2, and correct but incompletely resolved (with a trichotomy for taxa A, B, and C) for both locus 3 and for the combined loci. Thus the cladistic method was the only one of the three to accurately reconstruct phylogeny from the hypothetical data set, whether rare alleles were detected or not.

The qualitative coding, minimum turnover model (Mickevich & Mitter, 1981, 1983; Buth, 1984) can be outlined in six steps:

1. Treat each locus as a character.
2. Delete alleles found in only one taxon (autapomorphies) (although it is not clear whether Mickevich and Mitter are strong advocates of this particular practice).
3. Treat each allelic combination as a character state, ignoring frequencies (qualitative coding).
4. Order character states into character-state trees that minimize the total number of allelic changes (minimum-turnover model).
5. Root character-state trees by outgroup comparison.
6. Combine all character-state trees using the principle of parsimony (a computer program is usually required).

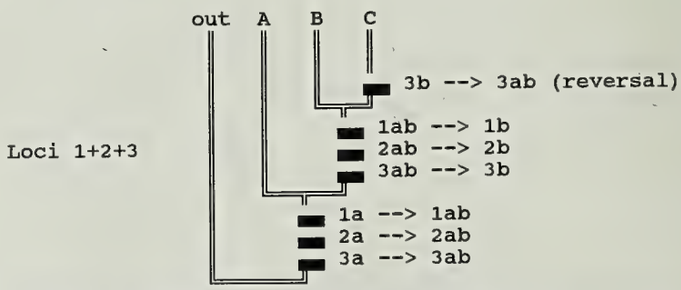


Figure 6. Application of cladistic step 6 (see text) to character-state trees of Figures 4 and 5. See text for explanation.

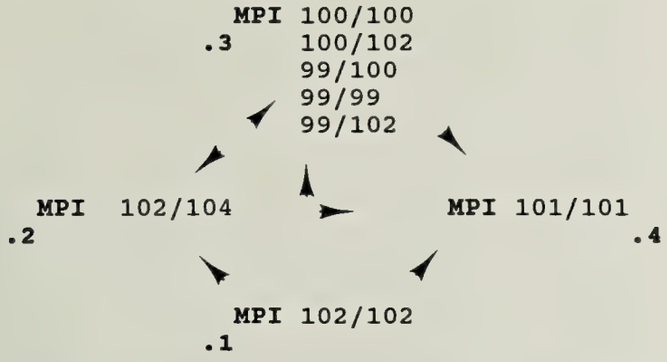


Figure 7. Detected allelic character states in the MPI locus among 20 genera of polygyrid snails, with hypothesized character-state assignments and transformations (from Emberton, in review).

These steps are demonstrated for the hypothetical phylogeny in Figures 4–6. Figure 4 shows loci 1 and 2, each of which yields the same cladogram, and each of which is treated separately (step 1). Since allele c occurs in only one taxon, it is deleted as a phylogenetically uninformative autapomorphy (step 2). This leaves three allelic combinations (a, b, and ab), each of which is treated as a single character state, regardless of allelic frequencies (step 3). Thus taxa B and C are scored equally for character-state b, even though this allele occurs at frequency 1.00 in taxon B and at frequency 0.05 in taxon C (Figure 4). Ordering these three character states by the minimum-turnover model (step 4) results in the order b <-> ab <-> a. This order is more parsimonious regarding allelic changes (i.e., turnover) than either alternative (i.e., b <-> a <-> ab or a <-> b <-> ab). Rooting this ordered character-state tree (a linear transformation series in this case) is done by outgroup comparison (step 5). The outgroup has character state a, which is therefore hypothesized as plesiomorphic. The rooted character-state tree, therefore, is a -> ab -> b for both locus 1 and locus 2. Each of these two loci has the same single-locus cladogram shown at the bottom of Figure 4.

Cladistic treatment of hypothetical locus 3 is similar and is outlined in Figure 5. The locus is treated independently (step 1); the autapomorphy for allele c in taxa A is deleted (step 2); the character states are defined as allelic combinations a, b, and ab (step 3), which are ordered and rooted in the tree a -> ab -> b (steps 3 and 4). The resulting single-locus cladogram (bottom of Figure 5) also puts taxa A, B, and C in a monophyletic clade defined by the synapomorphic transition a -> ab. This cladogram differs from those for loci 1 and 2 (Figure 4), however, in that it gives no further resolution: the transition ab -> b shows up only as an autapomorphy of taxon B.

Step 6 consists of combining the single-locus character-state trees or cladograms (Figures 4 and 5) in the most parsimonious way (i.e., involving the least amount of overall turnover among character states within their respective loci). This results in the cladogram of Figure 6, which is a perfectly accurate reconstruction of the original hypothetical phylogeny (Figure 1).

The preceding example was simple enough to be performed by hand. For complex data sets, the program Hennig86 (Farris, 1988) has been recommended (Platnick, 1989). For encoding branching character-state trees for Hennig86 analysis, binary coding is needed. The present author finds it useful to think of binary coding as a method of encoding not by character state but by transformation (between two character states). To encode the hypothetical data, each transformation is numbered as follows:

Locus	Transformation	Transformation Number
1	a → ab	0
1	ab → b	1
2	a → ab	2
2	ab → b	3
3	a → ab	4
3	ab → b	5

The following taxon-by-transformation matrix, then, is submitted to Hennig86. Each taxon is scored for the occurrence (1) or non-occurrence (0) of each transformation in the lineage that produced that taxon.

	Transformation					
Taxon	0	1	2	3	4	5
out	0	0	0	0	0	0
A	1	0	1	0	1	0
B	1	1	1	1	1	1
C	1	1	1	1	1	0

Given this matrix, Hennig86 produces the correct cladogram shown in Figure 6.

Applying this six-step method of allozyme cladistics to 20 polygyrid genera (Emberton, 1994) resulted in problems in steps 4 and 6. In step 4 (ordering allelic combinations into character-state trees), there was often a problem of a confusingly large number of allelic combinations, many of which occurred in complexly interrelated sets. For example, the mannose phosphate isomerase locus (MPI) yielded five alleles (99, 100, 101, 102, and 104, referring to their relative positions in mm on the electrophoretic gels). These alleles occurred in eight allelic combinations, which are shown in Figure 7. Ordering these eight combinations as individual character states would produce extreme complexity and, given the distributions of these combinations among taxa (Emberton, 1994), would yield apparently little additional phylogenetic information.

The ultimate solution to this problem seems to require writing a computer program to perform the minimum-turnover algorithm, regardless of the number of allelic combinations. An interim procedure, however, is to treat such interrelated sets as single character states. For the MPI example, this groups five interrelated allelic combinations as a single character state (Figure 7: character state .3). For other examples, see Emberton (1994).

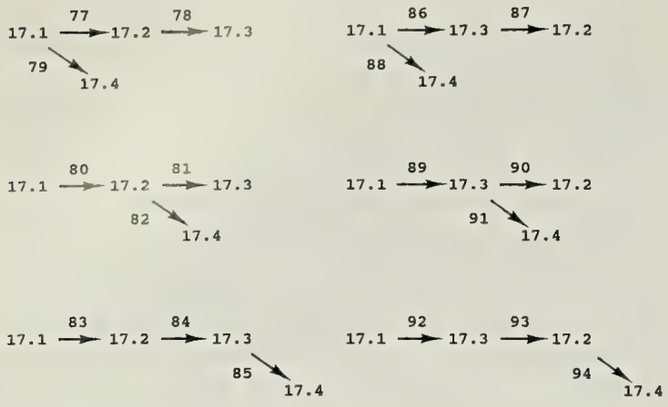


Figure 8. Alternative character-state trees derived from Figure 7. Numbering of character states and transformations as in Emberton (in review).

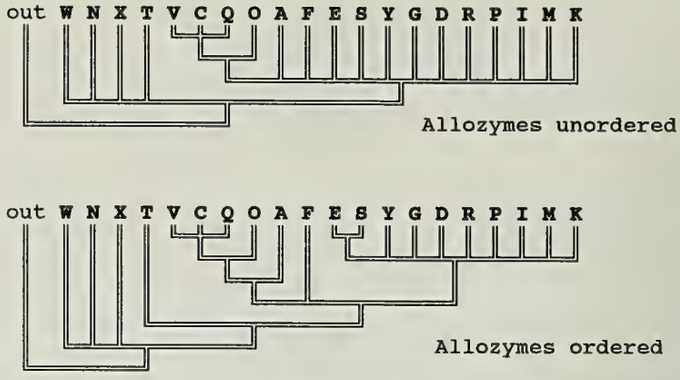


Figure 9. Cladograms from allozyme data on 20 polygyrid genera (from Emberton, in review).

Another problem that arose in applying step 6 (combining all character-state trees using parsimony) was that of alternative character-state trees. For example, in the MPI locus (Figure 7), state .3 has an equal probability (according to the minimum-turnover model: step 5) of being derived from states .1 or .2, but not from state .4. Likewise, state .4 has an equal probability of derivation from any of the three other states. To encode all these possible transformations would unduly weight information-poor loci over information-rich loci.

A solution devised by Emberton (1994) is to use all transformations in all alternative character-state trees, but to weight them by the reciprocal of the number of alternatives. For example, the multiple arrows among MPI character states .1, .2, .3, and .4 (Figure 7) result in six alternative character-state trees that are shown in Figure 8. Each of these trees was encoded for Hennig86 analysis (using the transformation-coding method described above, producing transformations numbers 77–94 as used in the complete analysis of polygyrid genera: Emberton, 1994). Before analysis, however, transformations 77–94 were all assigned a relative weight of 1/6 (using the ccode command of Hennig86). Transformations in other characters received other weights, depending on each character's total number of alternative character-state trees.

Another point needs to be made respecting step 4 (ordering allelic combinations into character-state trees). Although ordering of character states is considered by many cladists to be a central tenet of phylogenetic systematics (Hennig, 1966; Brooks & McLennan, 1991; Harvey & Pagel, 1991; Wilkinson, 1992), Hauser & Presch (1991; Presch, 1992) and others maintain that, at the very least, ordered data should also be analyzed unordered to determine the robustness of the cladogram to hypotheses of character-state order. This procedure applied to the polygyrid-genera allozyme data set (Emberton, 1994), yielded the comparison shown in Figure 9. The cladogram for ordered allozyme data has the same topology as that for unordered data, but with greater resolution (an increase from four nodes to eight nodes).

DISCUSSION

The present author's obvious preference for discrete-parsimony analysis of allozyme data (= allozyme cladistics) is mildly supported by the evaluation outlined in Figs. 1–3, which demonstrates a hypothetical case in which cladistics successfully reconstructs the true phylogeny–even when rare alleles are undetected–while both UPGMA and frequency parsimony (as interpreted and hand-calculated by the present author) fail. Further evaluations using both hypothetical and real data, and using other distance and parsimony methods, are needed to test these results, but it seems clear at least that the distance method of UPGMA bases phylogenetic inference on both plesiomorphic and apomorphic characters, including autapomorphies. On first principles, therefore, UPGMA (and, in the present author's incompletely informed opinion, other distance methods) should **not** be used to reconstruct phylogeny (Harvey & Pagel, 1991; Brooks & McLennan, 1991). Although it is true that UPGMA and cladistics results often are congruent, such cases only demonstrate relatively constant rates of allozyme evolution within the limitations of the data set, and do nothing to make the phylogenetic inference itself more robust.

Frequency parsimony (Swofford & Berlocher, 1987) depends heavily on frequencies, which can vary widely within a taxon (e.g. Emberton, 1993), and furthermore requires an algorithm that is very difficult to program and costly in computer time to run. A preliminary program is available (FREQPARS: Swofford, 1988; not used for the present paper), but can handle only very small data sets and often does not produce the most parsimonious solution(s) (Swofford, 1988; D. Lindberg, personal communication, 1992).

The practice of allozyme cladistics, as outlined above and demonstrated in Figs. 4–6, seems logical, objective, and empirically validated (Mickevich & Mitter, 1981, 1983; Buth, 1984). For polygyrid land-snail genera, it yielded a phylogenetic hypothesis that was generally both

consistent with and complementary to a hypothesis based on an independent anatomical data set (Emberton, 1994, unpublished).

Polymorphisms–both in allozyme and in morphological data– may at first seem a hindrance to cladistics. As Mickevich and Mitter's (1981, 1983) method points out, however, polymorphisms can offer important clues to character evolution, and hence to taxon evolution. Ordering of character states within characters may sometimes lead to error, however, so it is important to analyze data both ordered and unordered (Hauser & Presch, 1991; Wilkinson, 1992; Hauser, 1992). Ordering does enhance phylogenetic resolution without changing topology among polygyrid genera (Figure 9; Emberton, 1994), but does not among *Truncatella* snail species (G. Rosenberg, personal communication) and apparently does not among taxa in several non-molluscan data sets (Hauser & Presch, 1991). Clearly, each new data set must be evaluated in its own right.

ACKNOWLEDGMENTS

Supported in part by N.S.F. grant BSR–87–00198 to the author and N.I.H. grant TMPI11373 to G. M. Davis. I also thank Gary Rosenberg for helpful discussion, M.G. Harasewych for inviting my participation in this symposium, and Harasewych and two anonymous reviewers for useful comments on a previous draft.

LITERATURE CITED

Berger, E. M. 1983. Population genetics of marine gastropods and bivalves. *In:* Russell-Hunter, W. D. (ed.). The Mollusca, Volume 6, Ecology. Academic Press, New York, p. 563–596.

Brooks, D. R. and D. A. McLennan. 1991. Phylogeny, ecology and behavior: A research program in comparative biology. University of Chicago Press, Chicago, IL, 434 p.

Buth, D. G. 1984. The application of electrophoretic data in systematic studies. Annual Review of Ecology and Systematics 15:501–522.

Cain, A. J. 1983. Ecology and ecogenetics of terrestrial molluscan populations. *In:* Russell-Hunter, W. D. (ed.). The Mollusca, Volume 6, Ecology. Academic Press, New York, p. 597–647.

Davis, G. M., W. H. Heard, S. L. H. Fuller, and C. Hesterman. 1981. Molecular genetics and speciation in *Elliptio* and its relationships to other taxa of North American Unionidae (Bivalvia). Biological Journal of the Linnean Society (London) 15:131–150.

Emberton, K. C. 1988. The genitalic, allozymic, and conchological evolution of the eastern North American Triodopsinae (Gastropoda: Pulmonata: Polygyridae). Malacologia 28:159–273.

Emberton, K. C. 1991. The genitalic, allozymic, and conchological evolution of the tribe Mesodontini (Gastropoda: Pulmonata: Polygyridae). Malacologia 33:71–178.

Emberton, K. C. 1993. Over-representation of rare alleles in juveniles and lack of pattern in geographic distributions of alleles in a land snail. Malacologia 35:361–369.

Emberton, K. C. 1994. Polygyrid land-snail phylogeny: external sperm trading, early North American biogeography, iterative shell evolution. Biological Journal of the Linnean Society, 52: 1–31.

Farris, J. S. 1972. Estimating phylogenetic trees from distance matrices. American Naturalist 106:645–668.

Farris, J. S. 1988. Hennig86, Version 1.5. James S. Farris, Port Jefferson Station, New York, NY 11776.

Harvey, P. H. and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, New York, NY, p. 239.

Hauser, D. L. 1992. Similarity, falsification and character state order–a reply to Wilkinson. Cladistics 8:339–344.

Hauser, D. L. and W. Presch. 1991. The effect of ordered characters on phylogenetic reconstruction. Cladistics 7:243–265.

Hillis, D. M. 1989. Genetic consequences of partial self-fertilization on populations of *Liguus fasciatus* (Mollusca: Pulmonata: Bulimulidae). American Malacological Bulletin 7:7–12.

Hennig, W. 1966. Phylogenetic systematics. University of Illinois Press, Urbana, IL, p. 263.

Hillis, D. M. and C. Moritz (eds.). 1990. Molecular systematics. Sinauer Associates, Inc., Sunderland, MA, p. 588.

Hoeh, W. R. 1990. Phylogenetic relationships among eastern North American *Anodonta* (Bivalvia: Unionidae). Malacological Review 23:63–82.

Johnson, M. S., B. Clarke, and J. Murray. 1988. Discrepancies in the estimation of gene flow in *Partula*. Genetics 120: 233–238.

Johnson, M. S., J. Murray, and B. Clarke. 1986. An electrophoretic analysis of phylogeny and evolutionary rates in the genus *Partula* from the Society Islands. Proceedings of the Royal Society of London B 227:161–177.

Mickevich, M. F. and C. Mitter. 1981. Treating polymorphic characters in systematics: a phylogenetic treatment of electrophoretic data. *In:* Funk, V. A. and D. R. Brooks (eds.). Advances in cladistics. New York Botanical Garden, Bronx, NY, p. 45–60.

Mickevich, M. F. and C. Mitter. 1983. Evolutionary patterns in allozyme data: a systematic approach. *In:* Platnick, N. I. and V. A. Funk (eds.). Advances in cladistics 2. Columbia University Press, New York, NY, p. 169–176.

Mickevich, M. F. and S. J. Weller. 1990. Evolutionary character analysis: tracing character change on a cladogram. Cladistics 6:137–170.

Platnick, N. I. 1989. An empirical comparison of microcomputer parsimony programs, II. Cladistics 5:145–161.

Richardson, B. J., P. R. Baverstock, and M. Adams. 1986. Allozyme electrophoresis. A handbook for animal systematics and population studies. Academic Press, Sydney, p. 410.

Sarich, V. M. 1977. Rates, sample sizes, and the neutrality hypothesis for electrophoresis in evolutionary studies. Nature 265:24–28.

Sneath, P. H. A. and R. R. Sokal. 1973. Numerical taxonomy. Freeman, San Francisco, p. 573.

Swofford, D. L. 1988. FREQPARS, Release 1.0. David L. Swofford, Illinois State Natural History Survey, 172 Natural Resources Building, 607 E. Peabody, Champaign, IL 61820.

Swofford, D. L. and S. H. Berlocher. 1987. Inferring evolutionary trees from gene frequency data under the principle of maximum parsimony. Systematic Zoology 36:293–325.

Wiley, E. O. 1981. Phylogenetics. The theory and practice of phylogenetic systematics. John Wiley and Sons, New York, p. 439.

Wilkinson, M. 1992. Ordered versus unordered characters. Cladistics 8:375–385.

Woodruff, D. S. 1989. Genetic anomalies associated with *Cerion* hybrid zones: the origin and continuance of new electromorphic variants called hybrizymes. Biological Journal of the Linnean Society 36:281–294.

Woodruff, D. S. and A. Solem. 1990. Allozyme variation in the Australian camaenid land snail *Cristilabrum primum*: A prolegomenon for a molecular phylogeny of an extraordinary radiation in an isolated habitat. Veliger 33:129–139.

Wright, S. 1978. Evolution and the genetics of populations, Volume 4, Variability within and among natural populations. University of Chicago Press, Chicago, IL, 580 p.

Use of Random Amplified Polymorphic DNA (RAPD) Markers to Assess Relationships Among Beach Clams of the Genus *Donax*

S. Laura Adamkewicz
Department of Biology
George Mason University
Fairfax, VA 22030, USA

M. G. Harasewych
Department of Invertebrate Zoology
National Museum of Natural History
Smithsonian Institution
Washington, DC 20560, USA

ABSTRACT

The polymerase chain reaction was used to amplify genomic DNA from nine populations of donacid clams representing six taxa occurring in three sympatric pairs. The randomly amplified polymorphic DNA (RAPD) markers produced by this technique successfully distinguished among all taxa. Each taxon possessed a unique subset of markers and one member of each sympatric pair differed from the other by several markers. The taxa also separated clearly into two groups, one North American and the other Caribbean. Use of RAPD markers as characters in a cladistic analysis produced well resolved phylogenetic trees of high consistency.

Key words: RAPD, PCR, phylogeny, biogeography, *Donax*.

INTRODUCTION

Comparisons of closely related species are often informative with regard to the functional biology of their shared characters, while comparisons of sympatric congeners can provide insights to the selective forces and adaptive complexes that are important in speciation (Larson, 1989). Western Atlantic species of beach clams in the genus *Donax* are particularly appropriate candidates for such comparative studies. Six species or subspecies have been described from the coastal waters of the eastern United States, and nearly as many have been reported from the Caribbean (Morrison, 1971). All of these species are highly polymorphic for shell colors and patterns, and one (*Donax variabilis* Say, 1822) serves as a classic example of a hyper-variable species (Moment, 1962). Both in the western Atlantic and world-wide, *Donax* often occur as sympatric species pairs or triplets (Abbott, 1974; Ansell, 1983; Morrison, 1971) that partition their shared habitat. The selective forces that promote the rise and maintenance of hyper-variable polymorphisms are poorly understood (Allen, 1988; Owen & Whitely, 1988) and may be clarified by comparative studies. Furthermore, the interaction of habitat partitioning and hyper-variability has been studied only in the snail genus *Cepaea* (Clarke, 1960). Our long-term goal is to investigate these questions in *Donax*. However, the sympatric co-occurrences of two or more similar and often highly polymorphic species have placed the systematic status of some of these taxa in dispute.

Morrison (1971) recognized six taxa along the Atlantic and Gulf coasts of the United States: 1) *Donax fossor* Say, 1822, which ranges from New York to North Carolina; 2) *Donax variabilis variabilis* Say, 1822 (as *Donax roemeri protracta* Conrad, 1849[1]), which occurs from Virginia southward along both coasts of Florida and westward along the Gulf coast to Mississippi; 3) *Donax parvulus* Philippi, 1849, with a range that extends from North Carolina to the eastern coast of Florida; 4) *Donax dorotheae* Morrison, 1971, which occurs along the Gulf coast from Florida to Louisiana; 5) *Donax variabilis roemeri* Philippi, 1849 (as *Donax roemeri roemeri* Philippi, 1849), which ranges from the Mississippi delta westward along the coasts of Texas and Mexico; and 6) *Donax texasianus* Philippi, 1847 with the same range as *D. variabilis roemeri*. Only in the northern-most part of its American range is *Donax* represented by a single species, *D. fossor*. Elsewhere, species of *Donax* generally occur as sympatric pairs. In the Caribbean and along the coast of northern South America, the genus is represented by *Donax denticulatus denticulatus* Linné, 1758, *D. denticulatus stephaniae* Petuch, 1992, *D. striatus* Linné, 1767, and *D. vellicatus* Reeve, 1855, with two or three species occurring together. *Donax denticulatus* is the type species of the subgenus *Chion* Scopoli, 1777 (Gray, 1847).

Where species of *Donax* co-occur, they subdivide the habitat in much the same way. As described by Morrison (1971), *Donax parvulus*, *D. dorotheae*, and *D. texasianus* all occupy the same habitat and are allopatric, replacing one another along the coast, each co-occurring with *Donax variabilis*. In every case, *D. variabilis* is larger, occurs

[1] See Boss (1970) and Melville (1976) for details on the nomenclature of this taxon.

higher in the inter-tidal zone, and migrates more actively with the tides. The other three species are much smaller, occur at the bottom of the inter-tidal zone and spend much of the year sub-tidally. During at least some parts of the year, these species occur with *D. variabilis* and both can be collected in the same handful of sand. This co-occurrence, combined with morphological similarity, has placed the status and rank of several taxa in dispute. Abbott (1974) considered *D. parvulus* to be an offshore ecological form of *D. variabilis*. Loesch (1957) reviewed the names of donacid taxa reported from the Texas coast and stated that morphological intergrades had been reported between *D. texasianus* and what may have been *D. dorotheae* near Louisiana and between *D. texasianus* and *D. variabilis roemeri* along the Texas coast. Chanley (1969) suggested that *D. fossor* represents a [temporary, seasonal] summer extension of the range of *D. variabilis*.

In the Caribbean, *D. vellicatus*, like several of the American taxa, remains primarily sub-tidal. *Donax denticulatus* and *D. striatus* partition their habitat somewhat differently, but the division is still based on the tendency to migrate with the tide, as well as a preferred position in the inter-tidal zone. Wade (1967, 1968) has observed that, although *D. denticulatus* and *D. striatus* did sometimes occur together, *D. denticulatus* migrated actively throughout the tidal cycle while *D. striatus* maintained a constant position higher in the inter-tidal zone.

Historically, separations of donacid species and subspecies have been based exclusively on shell morphology (primarily on size, inflation [obesity], and degree of striation), which is subject to environmentally induced variation. Molecular data are often suitable for resolving taxonomic questions of this nature where the cause of morphological differences cannot be ascribed to either genetic or environmental differences. The restriction of a molecular character, either an allele in an allozyme system or a DNA marker, to one member of a sympatric species pair is taken as evidence of the absence of interbreeding between the two taxa. While molecular evidence cannot confirm genetic isolation when two populations are allopatric, it can at a minimum demonstrate genetic divergence. A previous attempt to differentiate between sympatric populations of *D. parvulus* and *D. variabilis* using allozyme data yielded ambiguous results (Nelson *et al.*, 1993). While allele frequencies differed between the two groups, no alleles unique to either taxon were found. Because the allozyme data could not distinguish the two taxa, additional molecular markers were sought.

The use of randomly amplified polymorphic DNA (RAPD) markers to differentiate between closely related individuals, populations and species was introduced in 1990 by Williams *et al.* and by Welsh and McClelland. In essence, segments of genomic DNA are amplified by the polymerase chain reaction (PCR) using a single very short primer (9 to 11 nucleotides) whose sequence might occur multiple times within the genome. The RAPD amplification of genomic DNA produces a set of fragments of various molecular weights, their number and size depending upon the number of times the primer sequence occurs in the genome, as well as the distances between pairs of primer sites. The RAPD technique has already been shown to produce genetic markers for Mendelian segregational analysis (Klein-Lankhorst *et al.*, 1991), genetic markers to distinguish among individuals within one population (Smith *et al.*, 1992), genetic markers that identify cultivars within a species (Hu & Quiros, 1991), and genetic markers that discriminate among species within a genus (Kambhampati *et al.*, 1992).

As a necessary prelude to the long-term goal of investigating morphological polymorphisms and ecological niche-partitioning in *Donax*, the present study seeks to resolve the systematic status of several western Atlantic donacid taxa and to discern their phylogenetic relationships. Failure of the allozyme data to resolve these taxa definitively has led us to investigate the utility of randomly amplified polymorphic DNA (RAPD) markers to distinguish between populations, subspecies and species in the genus *Donax*, as well as to determine the relationships of these populations and taxa using cladistic methodology.

MATERIALS AND METHODS

1. *COLLECTION OF SPECIMENS*

Donax were collected at the six locations shown on the map in Figure 7. These collections included samples from nine populations, representing six taxa, which are listed in Table 1 and are further identified as follows. 1) *Donax variabilis variabilis* (DVF) and 2) *Donax parvula* (DPF) were collected in the spring of 1990 at Indiatlantic Beach, on the Atlantic coast of Florida. 3) *Donax variabilis roemeri* (DVR) and 4) *Donax texasianus* (DTT) were collected at Corpus Christi, on the Gulf coast of Texas, in spring of 1990. 5) *Donax variabilis variabilis* (DVG) were collected in 1992 at Alligator Point on the Gulf coast of Florida. Collections were made in Jamaica in both May and November of 1991 for 6) *Donax denticulatus* (DDM) from Port Maria, a town on the northern coast of the island, along the town's seawall 7) *Donax denticulatus* (DDN) from Negril, a town on the western side of the island, at the Cosmos Beach Club 8) *Donax denticulatus* (DDB) and 9) *Donax striatus* (DSB) from Black River, a town on the southern coast of the island, at the Bridge House Inn.

Donax were collected by sieving sand from the intertidal zone of a beach. The animals were placed in plastic bags and kept cool until they could be identified to species, then frozen at −20° C and shipped to George Mason University. Thereafter, clams were maintained at −60° C until DNA was extracted. Attempts to collect *Donax fossor* and *D. dorotheae* at their respective type localities were unsuccessful, and these taxa are not included in the present study. Shells of the samples used in this study are deposited in the collections of the National Museum of Natural History, Smithsonian Institution. Catalogue numbers for voucher lots are listed in table 1.

Table 1. Sources of the nine populations of *Donax* used in this study. Collecting sites are those shown on the map in figure 7.

Sample designation	Taxon	Collecting site	USNM catalogue number
DVF	*Donax variabilis variabilis*	Atlantic coast of Florida	869538
DVG	*Donax variabilis variabilis*	Gulf coast of Florida	869539
DVR	*Donax variabilis roemeri*	Gulf coast of Texas	869540
DPF	*Donax parvulus*	Atlantic coast of Florida	869541
DTT	*Donax texasianus*	Gulf coast of Texas	869542
DDM	*Donax denticulatus*	Port Maria, Jamaica	869543
DDN	*Donax denticulatus*	Negril, Jamaica	869544
DDB	*Donax denticulatus*	Black River, Jamaica	869545
DSB	*Donax striatus*	Black River, Jamaica	869546

2. *EXTRACTION OF DNA*

To avoid contamination from food organisms, only muscle dissected from the foot of each clam was used to extract DNA. Approximately 30 mg of tissue was treated according to a protocol derived from that of Reeb and Avise (1990). Tissue was macerated in a 1.7 ml microcentrifuge tube containing 400 μl TE buffer (10 mM Tris, 1 mM EDTA, pH 7.6) for about 30 seconds with a pestle driven by an electric drill. After adding 25 μl of 10% sodium dodecyl sulfate (SDS), the extract was incubated at 65° C for 30 to 60 minutes. Next, 70 μl of 8M potassium acetate was added, the mixture shaken, and chilled on ice for 60 minutes. After the extract was centrifuged at 14,000 × g for 10 minutes, the supernatant was transferred to a clean tube and the pellet discarded. The supernatant was chilled at −20° C for 2 minutes, spun again for 10 minutes, and again transferred to a clean tube. Next, 400 μl of chloroform and 400 μl of tris-saturated phenol were added, the tube shaken, and centrifuged for 5 minutes at 14,000 × g. The upper, aqueous layer was transferred to a clean tube, 400 μl of chloroform added, the tube shaken, and centrifuged for 5 minutes. The upper, aqueous layer was again decanted to a clean tube, treated with chloroform, and centrifuged. After the upper layer was transferred to yet another clean tube, 1ml of cold 95% ethanol was added and mixed by gently inverting the tube. The sample was kept at −20°C for two minutes and centrifuged again for 10 minutes. The supernatant was discarded and the pellet was washed with 1 ml of 80% ethanol. After another centrifugation for 5 minutes at 14,000 × g, the ethanol was discarded and the pellet was dried in an incubator at 38° C for about 30 minutes. The cleaned DNA pellet was dissolved in 300 μl of TE buffer and kept at −20° C until needed.

This procedure yielded DNA at concentrations ranging from 5 to 35 μg/ml. If initial PCR amplification failed, the DNA was further purified with "GeneClean II" (BIO 101 Inc., P.O. Box 2284, La Jolla, CA 92038) after which it amplified satisfactorily.

3. *PCR AMPLIFICATION OF DNA*

The amplification protocol of Bowditch *et al.* (1993) was used in this study. "Amplitaq" DNA polymerase, supplied by Perkin-Elmer/Cetus at an activity of 8 units per μl, was used at a concentration of 0.5 units per sample (0.06 μl). The four nucleotide triphosphates were supplied by Pharmacia as 100mM stocks and mixed to make a single stock 0.25 mM for each dNTP. A special RAPD buffer was prepared according to the recipe: 100mM Tris, 500mM KCl, 19mM $MgCl_2$, and 10 mg/ml bovine serum albumin (not acetylated). Primers came from two sources: 20 (designated OP-E) were from Operon Technologies Kit E and 40 (designated LMS-P) were provided by the Laboratory for Molecular Systematics, National Museum of Natural History, Smithsonian Institution, where they had been synthesized.

Between 5 and 15ng of DNA from an individual clam and 50ng of a single, short primer (10mer in all cases) were combined in 25 μl of a reaction mixture comprised of: 18 μl sterile distilled water, 2.5 μl RAPD buffer, 2.5 μl deoxynucleotide mix, 1 μl primer, and 1 μl target DNA. The reaction mix was topped with mineral oil and placed in a Perkin-Elmer 4800 Thermocycler for 45 cycles of a RAPD amplification profile as follows: dissociation of DNA for 1 minute at 94°C, annealing of primer for 1 minute at 36°C, polymerization of DNA for 2 minutes at 72°C. The amplification products were loaded onto a 1.4% agarose gel, electrophoresed in TBE buffer (89mM Tris, 89mM Boric Acid, 2mM EDTA) and visualized with ethidium bromide. Each primer produced a characteristic set of amplification products with sizes ranging from 0.3 to 3.0 kilobases, which appeared as bright bands on the agarose gels (Figure 1). Rather than measuring distances from the origin, approximate sizes were determined by comparison to fragments of known size in a mixture of lambda DNA cut with HindIII and ϕX174 cut with HaeIII. To confirm that two bands were identical, samples were run in adjacent lanes of a gel. Throughout this paper, these products are referred to interchangeably as "amplification products," "DNA fragments," or "RAPD markers."

4. *SELECTION OF PRIMERS AND MARKERS*

Sixty primers were screened on a panel of 24 individuals comprised of three clams from each of the sample populations except *D. striatus* (DBS). Each primer was used

Figure 1. RAPD amplifications generated by primer OP-E18. The horizontal bars over the gel join the three individuals from each population that were run on each gel. Sample designations above the bars refer to taxa and populations listed in table 1. The letter M identifies lanes containing molecular weight standards (lambda DNA cut with HindIII + ϕX174 cut with HaeIII). Standard bands and sizes in kilobases are depicted to the right of the gel.

at least twice to amplify each screening DNA. A primer was judged to be suitable for use in this study if it met the criteria of: 1) amplification, that is, the production of clearly resolved DNA fragments, 2) reproducibility, with at least one DNA fragment appearing consistently and reproducibly in repeated assays of the same individuals, and 3) commonality, or the presence of at least one DNA fragment in two or more populations (but not necessarily two or more taxa). Primers that met these requirements were not common. Approximately one fourth of the primers tested failed to meet criterion 1, with most of the remainder failing criterion 3. Criterion 2, reproducibility, was not a serious problem. For all of the markers chosen, amplification of DNA from the same individual produced the same results whether the amplification was repeated in separate PCR experiments or replicated within the same PCR experiment. Identical results were produced when amplifications were repeated using a Coy thermocycler, in which temperature changes much more slowly than in a Perkin-Elmer machine.

The initial screening procedure identified five primers that produced a total of 17 RAPD markers that were informative for the purposes of this study. These primers and markers are described in Table 2 and representative results are shown in Figure 1. Of the two primers from the Laboratory for Molecular Systematics, primer LMS-P01 is the same as primer AP8g of Williams et al. (1990) while primer LMS-P56 was designed and synthesized at LMS. A total of nine individuals from each population were assayed at least twice with each of the five primers. Samples with similar markers were run side-by-side in the replicate assay in order to facilitate direct comparisons.

5. *PHYLOGENETIC ANALYSIS*

Data were analyzed and trees produced using Hennig86 version 1.5 software (Farris, 1988). The implicit enumeration (ie;) algorithm was used in each series of analyses to insure that all shortest, equally parsimonious trees were found.

Each RAPD marker was treated as a separate character regardless of which primer was used to generate it or which other bands from the same or other primers co-occurred with it. In an initial analysis (Analysis 1), the 17 markers listed in Table 4 were scored as either absent (0) or present (1) for each population, and a hypothetical outgroup, scored as lacking all 17 RAPD markers (all characters = 0), was used (for data matrix, see Appendix 1). In subsequent analyses, the markers were scored as absent (0), polymorphic (1), *i.e.* present in some but not all members of the population, or fixed (2), *i.e.* present in all members of the population. In Analysis 2, this data set was run unordered, using the same hypothetical outgroup as in Analysis 1 (data matrix in Appendix 2). A third series of analyses used the same data matrix but, instead of the hypothetical outgroup, each of the basal taxa from Analysis 2 (DSB, DDN) was used in turn as the outgroup. The arrangement of character states on the resulting trees were examined using the Dos Equis (*xx*) and *xsteps* tree diagnostic commands.

Table 2. DNA primers used in this study. For each primer the table shows: the sequence; the average number of DNA amplification products detected in an individual, with the range of averages among the nine groups following in parentheses; and the sizes of those DNA amplification products used as markers. Primers designated OP-E are from Operon Technologies Kit E and primers designated LMS-P were provided by the Laboratory for Molecular Systematics. Those fragments followed by an asterisk (*) are characteristic of the Caribbean taxa, while those followed by an ampersand (&) are characteristic of the Carolinian taxa. Note that, for any given primer, the smaller fragments are always characteristic of the Carolinian taxa.

Primer	Sequence	Average number of RAPD DNA fragments per clam (range)	Size (kb) of useful RAPD markers
OP-E07	5′-AGATGCAGCC	3.8 (3.1–4.3)	0.6 1.5
OP-E16	5′-GGTGACTGTG	3.4 (2.6–4.3)	0.3& 0.5& 0.6* 0.9 1.1
OP-E18	5′-GGACTGCAGA	3.0 (1.6–4.2)	0.5& 0.6* 0.9
LMS-P01	5′-TGGTCAGTGA	2.8 (2.0–3.2)	0.5& 1.0* 1.2*
LMS-P56	5′-AGATCTGCAG	3.0 (2.2–3.6)	0.3& 0.6& 1.1* 1.2*

Table 3. Distribution of RAPD DNA markers in May and November samples of *Donax denticulatus* from Port Maria, Jamaica. As in tables 2 and 4, the RAPD marker identification designates the primer that produced the marker, the approximate size of each marker in kilobases, and our identifying marker number. Each entry in the matrix shows the number of individuals in which the RAPD marker was detected over the number of individuals tested (*e.g.*, 3/9). Because Primer OP-E07 was not used with the May sample, data from this sample were not included in table 4, which serves as the data matrix for phylogenetic analysis.

Primer:	OP E16	OP E16	OP E18	LMS P01	LMS P56	LMS P56	OP E16	OP E18	LMS P01
Marker size (kb):	0.9	1.1	0.9	1.2	1.1	1.2	0.6	0.6	1.0
Marker number:	7	9	10	12	13	14	15	16	17
May	9/9	9/9	9/9	0/9	4/8	8/8	9/9	9/9	9/9
November	9/9	9/9	9/9	1/9	6/6	5/8	8/9	9/9	8/9

RESULTS

1. *STABILITY OF RAPD MARKERS OVER TIME*

To assess the stability of the RAPD markers used in this study, two different samples of *Donax denticulatus* from Port Maria, Jamaica, were examined, one taken in May and the other in November of 1991. These two samples produced identical RAPD markers with only slight differences in frequencies between the two collections (Table 3). The only marker not present in both samples, marker 12 (LMS-P01/1.2kb), was the rarest, appearing in only 3 of a total of 21 individuals assayed.

2. *DISTRIBUTION OF RAPD MARKERS AMONG POPULATIONS*

Table 4 summarizes the distribution of the 17 RAPD markers among the nine assayed populations, but does not contain data from the May sample of *Donax denticulatus* from Port Maria. Because the presence of two species in the sample from Black River was not discovered until after the laboratory work was completed, these two populations have reduced sample sizes (3 *D. denticulatus*, 6 *D. striatus*).

Of the 17 RAPD markers assayed, two (markers 8 and 9), each produced by a different primer, were present at varying frequencies in all nine populations. All three samples of *D. variabilis* (DVF, DVG, DVR), including the subspecies *D. variabilis roemeri*, had the same 11 markers appearing in at least one member of each population. Of these 11 markers, two appeared in no other taxon. Similarly, all three populations of *D. denticulatus* (DDM, DDN, DDB) shared a set of 11 markers, three of which appeared in no other taxon. Twelve of the 17 RAPD markers were not shared between *D. variabilis* (DVF, DVG,DVR) and *D. denticulatus* (DDM, DDN, DDB). The three remaining taxa, each represented by a single population, showed clear affinities with either *D. variabilis* or *D. denticulatus*. The absence of unique markers in these taxa was an artifact of the criteria for primer selection (i.e.- that bands occur in at least two sample populations). Nevertheless, *Donax parvulus* (DPF) was distinguished from its sympatric congener *D. variabilis variabilis* (DVF) by the absence of RAPD markers 1 and 2, while *D. texasianus* (DTT) differed from its sympatric congener *D. variabilis roemeri* (DVR) in lacking markers 1, 2, 3, and 7. Marker 2, which was fixed in all populations of *Donax variabilis* and present in no other taxon, appears to be a diagnostic marker for this species. The Caribbean species *D. striatus* (DSB) differed from *D. denticulatus* (DDB) in lacking 5 RAPD markers (10, 11, 15, 16, 17), of which two (10, 16) were fixed in *D. denticulatus*. Marker 16, which occurred in all individuals of *D. denticulatus* tested, was unique to this species and may be used as a diagnostic marker for this species.

Although diagnostic markers were not identified for some taxa, the absence of multiple RAPD markers in one member of each sympatric pair (including markers fixed in the other member of the sympatric pair) is taken as evidence that these pairs do not exchange genetic material. An empirical observation is that when primers produced RAPD markers characteristic of both Carolinian and Caribbean taxa, markers that distinguished Carolinian taxa were invariably shorter than markers that were diagnostic of Caribbean taxa (Table 2). The significance of this observation is not yet clear.

3. *PHYLOGENETIC ANALYSES*

An initial cladistic analysis (Analysis 1), scoring each of the 17 RAPD markers as absent or present (regardless of frequency) in each sample population and using a hypothetical outgroup in which all characters were scored as absent, produced a single most parsimonious tree (length = 20, ci = 85, ri = 91) that resolved all species level taxa but left populations and/or subspecies unresolved (Figure 2). Fourteen of the 17 character transformations plotted unambiguously onto this tree (Figure 2). Each of the remaining three characters (markers 7, 10, 11) could either have been present in the common ancestor of all the taxa in this study and subsequently lost in a single taxon (Figure 2), or have arisen twice independently (Figure 3). An analysis of character polarity using the out-group comparison method (Watrous & Wheeler, 1981) indicated that the presence of markers 7, 10, and 11 is plesiomorphic, as they occur in both the in-group and the out-group, while their loss, in each case

Table 4. Distribution of the RAPD DNA markers among sample populations. For each RAPD marker, the band designation identifies the primer, the approximate size of the marker in kilobases and a sequential number to identify the marker. Each entry in the matrix shows the number of individuals in which the RAPD marker was detected over the number of individuals tested (*e.g.*, 3/9). In a few cases, the number of individuals tested was less than nine, either because fewer DNA samples were available (DDB, DSB) or because we were unable to score an individual for a particular marker (markers 1, 13, 14). When no individuals produced a marker, the entry is marked "—" rather than 0/n. Data are grouped to emphasize affinities among sample populations, rather than by primer.

Primer:	OP	OP	LMS	OP	LMS	LMS	OP	OP	OP	OP	LMS	LMS	LMS	OP	OP	OP	LMS
Size (kb):	E16	E18	P56	E16	P01	P56	E16	E07	E16	E18	E07	P01	P56	E16	E18	E16	P01
Number:	0.3	0.5	0.3	0.5	0.5	0.6	0.9	1.5	1.1	0.9	0.6	1.2	1.1	0.6	0.6	0.6	1.0
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
DVF	2/9	9/9	8/9	9/9	9/9	9/9	5/9	9/9	3/9	9/9	9/9	—	—	—	—	—	—
DVG	1/9	9/9	9/9	9/9	9/9	9/9	5/9	9/9	4/9	9/9	9/9	—	—	—	—	—	—
DVR	4/8	9/9	6/9	9/9	9/9	9/9	5/9	9/9	6/9	9/9	7/9	—	—	—	—	—	—
DPF	—	—	5/9	7/9	9/9	9/9	7/9	6/9	2/9	9/9	2/9	—	—	—	—	—	—
DTT	—	—	—	9/9	3/9	9/9	—	9/9	9/9	2/9	2/9	—	—	—	—	—	—
DDM	—	—	—	—	—	—	9/9	2/9	9/9	9/9	5/9	1/9	6/6	5/8	8/9	9/9	8/9
DDN	—	—	—	—	—	—	9/9	7/9	9/9	9/9	7/9	1/9	5/9	6/9	8/9	9/9	9/9
DDB[1]	—	—	—	—	—	—	3/3	2/3	3/3	3/3	2/3	1/3	3/3	1/3	2/3	3/3	3/3
DSB[2]	—	—	—	—	—	—	6/6	5/6	6/6	—	—	6/6	4/6	6/6	—	—	—

[1] Only 3 individuals available for testing.
[2] Only 6 individuals available for testing.

restricted to a functional in-group, is apomorphic (Fig 2).

A second analysis, scoring the RAPD markers as absent (0), polymorphic (1) or fixed (2), and employing the same hypothetical outgroup, produced four equally parsimonious trees (length = 32, ci = 90, ri = 92) when the data were run unordered. One tree, supported by all markers except 8, 10, and 11, matched the topology of the tree in Figure 3, except that all sample populations were resolved. In the other three trees, which differed only in the resolution of *D. denticulatus* populations and which were supported by all markers except 7, 8, and 13, *Donax striatus* emerged as the sister group to all remaining taxa. The nelsen consensus tree (length = 34, ci = 85, ri = 87) of these four trees is shown in Figure 4. Twelve of the 17 character transformations plotted uniquely onto the consensus tree, while markers (7, 8, 10, 11, 13) could be interpreted as evolving in several equally parsimonious scenarios. Analyses of character polarity using the outgroup comparison method (Watrous & Wheeler, 1981) suggest that marker 7 was fixed in the *Donax* ancestor, became polymorphic in the North American clade, and eventually lost in *Donax texasianus*, while marker 11, which was polymorphic in the *Donax* ancestor, was lost in *Donax striatus*, but became fixed in Florida populations of *Donax variabilis*. The remaining markers could not be mapped onto the consensus tree (Figure 4) without reversals (markers 8, 13) or convergences (marker 10). Only marker 8 was incompatible with all of the initial trees.

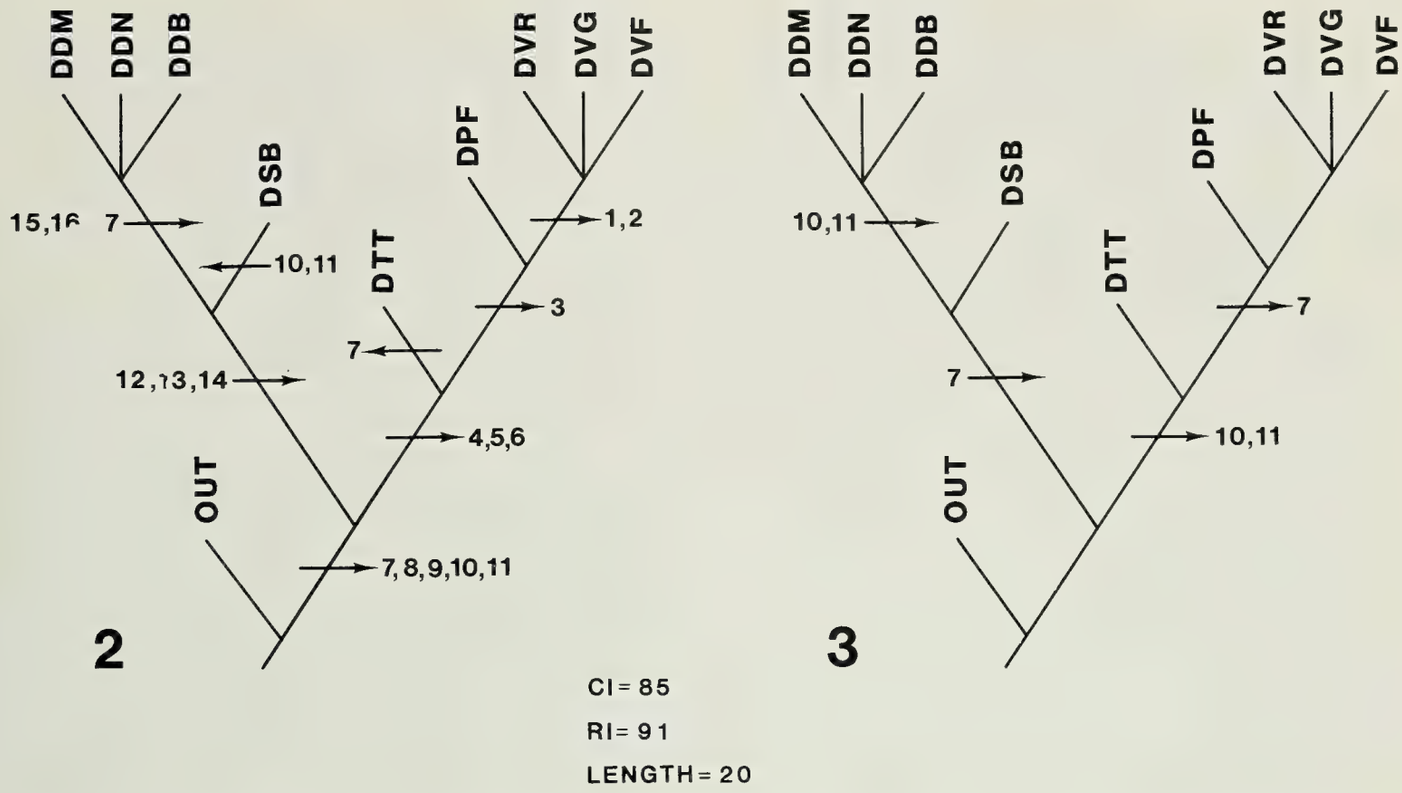
When the data were reanalyzed using *Donax striatus* as the outgroup, one most parsimonious tree (length = 28, ci =96, ri = 97), resulted (Figure 5). Likewise, when the Black River population of *Donax denticulatus*, served as the outgroup, a single, equally parsimonious tree (length = 28, ci =96, ri = 97), was produced (Figure 6).

All analyses that employed an intermediate character state produced identical tree topologies for the Carolinian samples but differed in the resolution and/or relationships of the Caribbean taxa and populations.

DISCUSSION

RAPD markers observed in the nine samples, representing six species or subspecies of *Donax*, showed a high degree of polymorphism both within and among taxa. Nevertheless, the polymorphisms did not obscure relationships among the samples and the presence of these markers was stable over time. The distribution of these RAPD markers supports previously disputed distinctions between members of the following three sympatric pairs of species: *Donax parvulus* and *D. variabilis variabilis*, *D. texasianus* and *D. variabilis roemeri*, and *D. denticulatus* and *D. striatus*, as well as between the similar but allopatric pair *D. parvulus* and *D. texasianus*.

Among the earlier applications of RAPD methodology, the technique was used to distinguish between individuals, strains, cultivars, populations and species (e.g. Hu & Quiros, 1991, Kambhampati et al. 1992). The char-



Figures 2–3. Phylogenetic tree resulting from analysis 1, in which all RAPD markers listed in table 4 were scored as absent or present in each sample population, regardless of frequency (data matrix in appendix 1). **2.** Character transformations plotted onto tree. Markers 7, 10 and 11 are plotted as having arisen once and been subsequently lost in a single taxon. **3.** Markers 7, 10 and 11 are plotted as having arisen twice independently.

acterization of diagnostic markers that would identify *Donax* species was beyond the scope of this study, and our criteria for primer selection precluded the recognition of markers unique to taxa represented by single samples. Even with this screening bias, fixed, species specific RAPD markers were discovered for *Donax variabilis* s.l. and for *D. denticulatus*, the two species in this study that were represented by multiple samples.

Kambhampati *et al.* (1992) successfully applied RAPD methodology to identify mosquito species and were able to cluster individuals of the same species correctly by applying phenetic algorithms (UPGMA) to markers generated by two primers. Although the resulting phenogram did not reflect the ancestral relationships of the mosquito species, these authors did not rule out the utility of RAPD data for phylogeny reconstruction but suggested that a greater number of primers (>20) should be tested to find lineage-specific markers. Our results confirm their conjecture. A survey of 60 primers was necessary to select the 5 primers and 17 markers that we

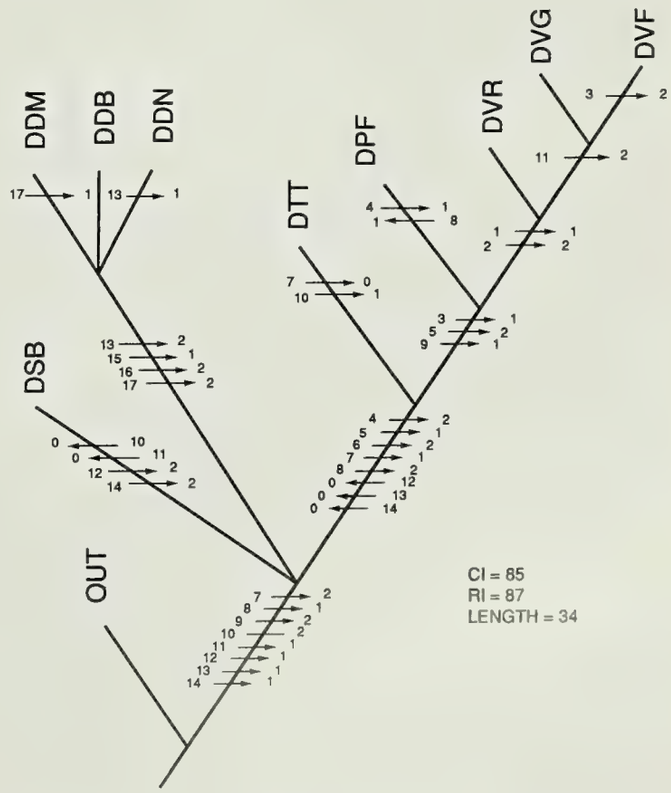


Figure 4. Nelson Consensus Tree of four equally parsimonious trees produced when all RAPD markers listed in table 4 were scored as absent, polymorphic, or fixed (data matrix in appendix 2) and all characters were run unordered. Character transformations are plotted onto the tree. Where alternative, equally parsimonious character transformations were possible, preference was given to the ordered, but not polarized, transformation series (0 <-> 1 <-> 2) and to reversals (loss of marker) over convergent origins of a marker.

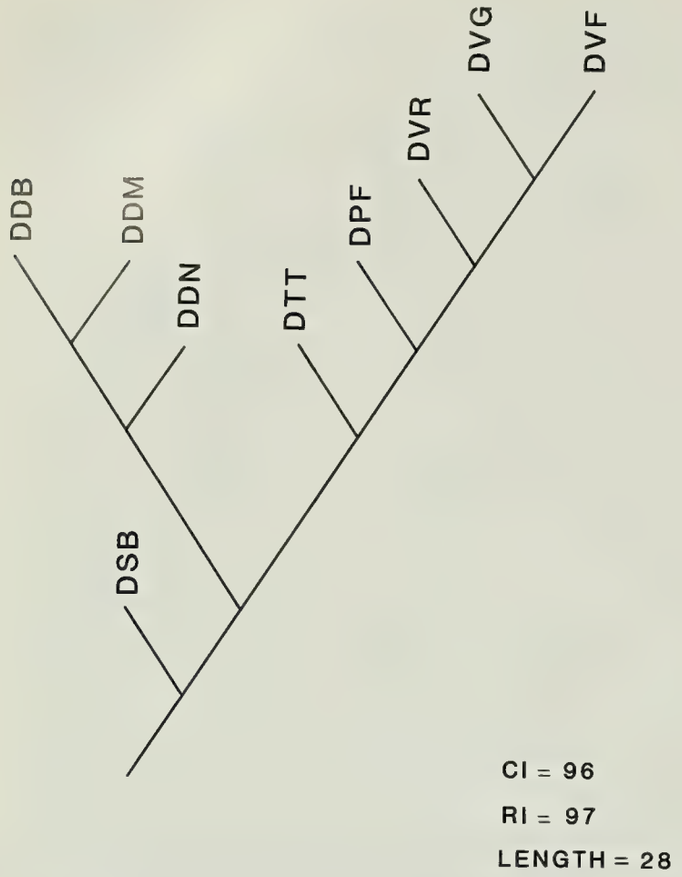


Figure 5. Most parsimonious tree produced when all RAPD markers listed in table 4 were scored as absent, polymorphic, or fixed, and all characters were run unordered. Based upon results of the previous analysis (Figure 4), *Donax striatus* was selected as the outgroup.

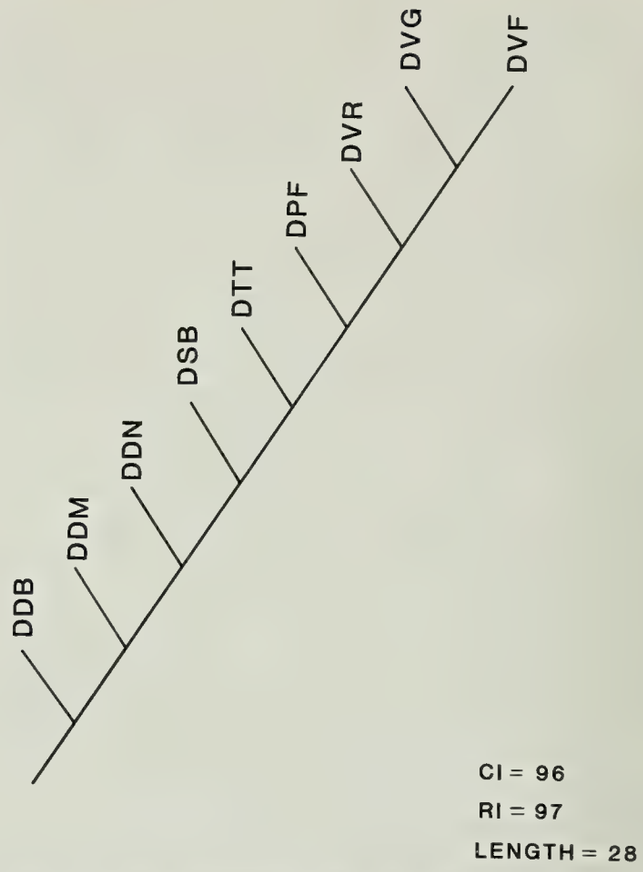


Figure 6. Most parsimonious tree produced when all RAPD markers listed in table 4 were scored as absent, polymorphic, or fixed, and all characters were run unordered. Based upon results of a previous analysis (Figure 4), the Black River population *Donax denticulatus*, was selected as the outgroup.

used to construct phylogenies. As RAPD markers have previously been shown to segregate in a Mendelian manner, behaving as dominant alleles (Williams *et al.*, 1990), we treated the individual markers as homologous characters suitable for cladistic analyses. Although the analyses were run unordered, when character transformations could be plotted onto the resulting tree in several equally parsimonious ways, preference was given to the ordered, but not polarized, transformation series (0 <-> 1 <-> 2) because this series reflects the manner in which alleles enter, are distributed within, and leave populations. Preference was also given to reversal (loss of a marker) over convergent evolution of a marker. Despite differences in outgroup selection and data scoring, all cladistic analyses produced the same, single, highly consistent tree for the five Carolinian samples. The inability to stably resolve the Caribbean samples is attributed in large part to the low number of samples.

A plot of the consensus tree (Figure 4) on a map of geographic distributions of the taxa (Figure 7) illustrates that the Carolinian *Donax* species form a monophyletic clade, while the Caribbean species may represent either a clade (Figure 2) or grade (Figs. 5,6) depending on how the limited data are analyzed. Although the genus *Donax*

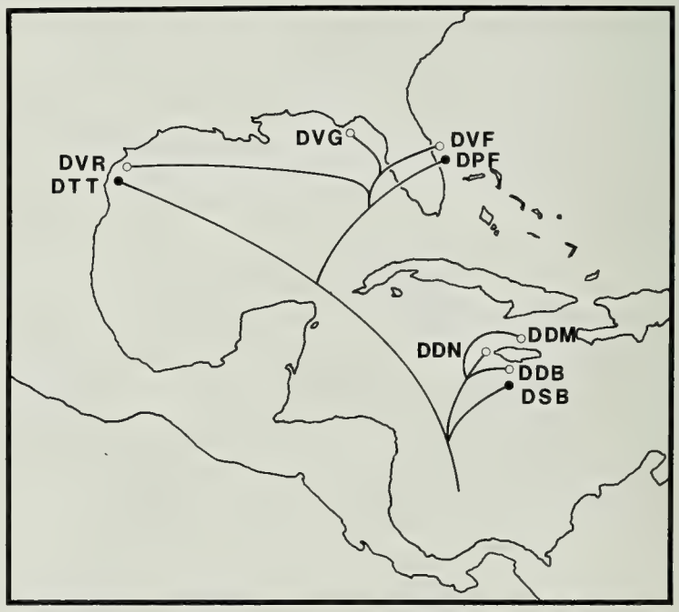


Figure 7. A plot of the consensus tree (Figure 4) on a map of the locations of the collection sites for the nine sample populations of *Donax* used in this study. Open circles indicate species that migrate with the tide. Solid circles indicate subtidal species. Sample designations as in Table 1.

has been represented in the fossil record of the western Atlantic since the Oligocene (Gardner, 1943:105), provincial boundaries have existed between molluscan faunas of the Gulf of Mexico and the Caribbean Sea since the late Oligocene or early Miocene (Petuch, 1988:48). The Recent Carolinian and Caribbean Provinces comprise, respectively, the Caloosahatchian Province and a portion of the larger Gatunian Province, both ranging from the late Oligocene to the early Pleistocene (Petuch, 1988:fig.1). Thus, the considerable divergence between Carolinian and Caribbean *Donax* faunas, which share at most five of 17 markers (29% similarity), may have accumulated over a period of approximately 25 million years.

The topology of the phylogenetic tree indicates that, within the Carolinian province, the non-migratory, lower intertidal to subtidal habitat is the more primitive among *Donax*. Species occupying this habitat (*D. texasianus*, *D. parvulus* and *D. dorotheae*) likely diverged as a result of barriers to gene flow posed by the Mississippi River and the emergence of peninsular Florida, respectively. As *D. parvulus* appears to be the sister species of *D. variabilis*, it is likely that *D. variabilis* originated in the eastern Carolinian Province, although comparable RAPD data on *D. dorotheae* may make possible a more precise localization of the area of origin of *D. variabilis*.

ACKNOWLEDGMENTS

The authors are indebted to the staff of the Laboratory for Molecular Systematics (LMS), National Museum of Natural History, Smithsonian Institution, and particularly to Darrilyn Albright, for technical advice. The authors learned the RAPD technique at LMS and, with the help of John Slapcinsky, now at the Field Museum of Natural History, generated all of the RAPD data while working there. George Mason University provided major support for this project in the form of a sabbatical leave for S.L. Adamkewicz while a grant from Jeffress Trust provided her financial support. The authors thank Walter Nelson, Florida Institute of Technology, Paul Mikkelsen, Palm Beach County Department of Environmental Resource management, and Robert Vega, Texas Parks and Wildlife Department, who collected some of the samples. We are grateful to Prof. Diana Lipscomb and Mr. John B. Wise of George Washington University for helpful discussions on phylogeny reconstruction. S.L. Adamkewicz also wishes to thank her husband for his untiring help in collecting specimens for this project and for all earlier ones.

LITERATURE CITED

Abbott, R.T. 1974. American Seashells, Second Edition. Van Nostrand, Reinhold Company, New York. 663pp.

Allen, J.A. 1988. Reflexive selection is apostatic selection. Oikos 51:251–252.

Ansell, A.D. 1983. The biology of the genus *Donax*. *IN* Developments in Hydrobiology 19, sandy beaches as ecosystems, pp. 607–634. A. McLachlan and T. Erasmus editors. H.J. Dumont, series editor. Dr. W. Junk Publishers. The Hague.

Boss, K. J. 1970. *Donax variabilis* Say, 1822 (Mollusca: Bivalvia): proposed validation under the plenary powers. Z. N. (S.) 1923. Bulletin of Zoological Nomenclature 27:205–206.

Bowditch, B.M., D.G. Albright, J.G.K. Williams and M.J. Braun. 1993. The use of RAPD markers in comparative genome studies. Methods in Enzymology 224:294–309.

Chanley, P. 1969. *Donax fossor*: a summer range extension of *Donax variabilis*. The Nautilus 83(1):1–14.

Clarke, B. 1960. Divergent effects of natural selection on two closely-related polymorphic snails. Heredity 14:423–443.

Farris, J. S. 1988. Hennig86 Reference. Version 1.5 (18 pp. manual distributed with program diskette).

Gardner, J. 1943. Mollusca from the Miocene and Lower Pliocene of Virginia and North Carolina. Part I. Pelecypoda. United States Geological Survey Professional Paper 199-A. 178 p. 23 pls.

Gray, J. E. 1847. A list of the Genera of the Recent Mollusca, their synonyma and types. Proceedings of the Zoological Society of London 1847:129–219.

Hu, J. and C. F. Quiros. 1991. Identification of broccoli and cauliflower cultivars with RAPD markers. Plant Cell Reports 10:505–511.

Kambhampati, S., W.C. Black IV and K. S. Rai. 1992. Random Amplified Polymorphic DNA of Mosquito Species and Populations (Diptera: Culicidae): Techniques, Statistical Analysis, and Applications. Journal of Medical Entomology 29(6): 939–945.

Klein-Lankhorst, R.M., A. Vermunt, R. Weide, T. Liharska, and P. Zabel. 1991. Isolation of molecular markers for tomato (*L. esculentum*) using random amplified polymorphic DNA (RAPD). Theoretical and Applied Genetics 83:108–114.

Larson, A. 1989. The relationship between speciation and morphological evolution. *In*: Otte, D. and J. A. Endler (eds.). Speciation and Its Consequences. Sinauer Associates, Inc. Sunderland, MA. pp. 579–598.

Loesch, H.C. 1957. Studies on the ecology of two species of *Donax* on Mustang Island, Texas. Publications of the Institute of Marine Science, University of Texas 4:201–227.

Mayr, E. 1963. Animal Species and Evolution. Harvard University Press, Cambridge, MA. 797 pp.

Melville, R. V. 1976. Opinion 1057. *Donax variabilis* Schumacher, 1817) Mollusca: Bivalvia) suppressed under the plenary powers; type species designated for *Latona* Schumacher, 1817. Bulletin of Zoolological Nomenclature 33(1): 19–21.

Moment, G.B. 1962. Reflexive selection: a possible answer to an old puzzle. Science 136:262–263.

Morrison, J. P. E. 1971. Western Atlantic *Donax*. Proceedings of the Biological Society of Washington 83(48): 545–568, 2 pls.

Nelson, W.G., E. Bonsdorff, and L. Adamkewicz. 1993. Ecological, morphological, and genetic differences between the sympatric bivalves *Donax variabilis* Say, 1822, and *Donax parvula* Phillipi, 1849. The Veliger 36(4):317–322.

Owen, D. F. and D. Whiteley. 1988. The beach clams of Thessalonika: reflexive or apostatic selection? Oikos 51: 253–255.

Petuch, E. J. 1988. Neogene history of tropical American mollusks, biogeography and evolutionary patterns of tropical western Atlantic Mollusca. The Coastal Education and Research Foundation, Charlottesville, 217pp.

Reeb, C.A. and J.C. Avise. 1990. A genetic discontinuity in a continuously distributed species: Mitochondrial DNA in the American Oyster, *Crassostrea virginica*. Genetics 124: 397–406.

Smith, M.L., J.N. Bruhn, and J.B. Anderson. 1992. The fungus *Armillaria bulbosa* is among the largest and oldest living organisms. Nature 356:428–431.

Wade, B. A. 1967. Studies on the West Indian beach clam, *Donax denticulatus* Linne. 1. Ecology. Bulletin of Marine Science 17:149–174.

Wade, B. A. 1968. Studies on the West Indian beach clam, *Donax denticulatus* Linne. 2. Life History. Bulletin of Marine Science 18:877–901.

Watrous, L. E. and Q. D. Wheeler. 1981. The out-group comparison method for character analysis. Systematic Zoology 30(1):1–11.

Welsh, J. and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acid Research 18:7213–7218.

Williams, J. G. K., Kubelik, A. E., Levak, K. J., Rafalski, J. A. and Tingey, S. C. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acid Research 18:6531–6535.

Appendix 1. Data matrix for initial cladistic analysis of *Donax* phylogeny. RAPD markers scored as absent (0) or present (1), regardless of frequency. For sample designations see Table 1.

	RAPD markers																
Samples	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Outgroup	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DVF	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
DVG	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
DVR	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
DPF	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
DTT	0	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0
DDM	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
DDN	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
DDB	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
DSB	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	0

Appendix 2. Data matrix for subsequent cladistic analyses of *Donax* phylogeny. RAPD markers scored as absent (0), polymorphic (1), or fixed (2). For sample designations see Table 1.

	RAPD markers																
Samples	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Outgroup	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DVF	1	2	1	2	2	2	1	2	1	2	2	0	0	0	0	0	0
DVG	1	2	2	2	2	2	1	2	1	2	2	0	0	0	0	0	0
DVR	1	2	1	2	2	2	1	2	1	2	1	0	0	0	0	0	0
DPF	0	0	1	1	2	2	1	1	1	2	1	0	0	0	0	0	0
DTT	0	0	0	2	1	2	0	2	2	1	1	0	0	0	0	0	0
DDM	0	0	0	0	0	0	2	1	2	2	1	1	2	1	1	2	1
DDN	0	0	0	0	0	0	2	1	2	2	1	1	1	1	1	2	2
DDB	0	0	0	0	0	0	2	1	2	2	1	1	2	1	1	2	2
DSB	0	0	0	0	0	0	2	1	2	0	0	2	1	2	0	0	0

Mitochondrial Genomes and the Phylogeny of Mollusks

Jeffrey L. Boore[1]
Wesley M. Brown
Department of Biology
University of Michigan
830 N. University Ave.
Ann Arbor, MI 48109-1048, USA

ABSTRACT

We are seeking a character set that reliably reflects the evolutionary origin of the phylum Mollusca and the relationships among molluscan classes. Such a character set must be: (1) present in all taxa; (2) unambiguously homologous; (3) changing at a rate appropriate for the taxonomic range; and (4) complex enough to make convergence highly unlikely. The arrangement of genes in mitochondrial DNA (mtDNA) appears to meet these criteria. With only a few exceptions, the mtDNA of all metazoans contains the same 37 genes: 2 for ribosomal RNAs, 22 for transfer RNAs, and 13 for proteins. Comparing the arrangements of these genes among the 17 taxa for which they are known suggests that the rate of change is appropriate for resolving higher level relationships. These genes could potentially be arranged in $>2 \times 10^{52}$ different ways; thus, the probability of the same order by convergence is very small. We have determined the complete mtDNA sequence of the bivalve *Mytilus edulis* (Hoffmann, Boore & Brown, 1992). This mtDNA differs from those of other metazoans in its unique gene arrangement, in encoding an additional tRNA (tRNAmet(AUA)), and in lacking one protein coding gene (ATPase 8). In order to test whether these features are typical for mollusks, and to investigate mitochondrial gene arrangements as a phylogenetic character for molluscan relationships, we have determined the mtDNA sequence for the polyplacophoran *Katharina tunicata*. *Katharina* mtDNA contains the gene for ATPase 8 and has a gene arrangement substantially different from that of *Mytilus* and much more similar to that of *Drosophila*. The different gene arrangements of *Mytilus* and *Katharina* provide numerous character states for investigating molluscan class relationships. By screening for gene junctions unique to one of these two arrangements, it may be possible to find patterns of rearrangements which unite the remaining classes to reflect their evolutionary history. The DNA sequence already obtained allows rapid screening of other animals by two methods. First, the polymerase chain reaction (PCR) can be used to selectively amplify gene boundaries unique to one of these two arrangements. Second, a large number of animals can be rapidly tested for the general arrangement of several of the large, well-conserved genes by Southern blot analysis. We compare the mitochondrial genome arrangements of *Mytilus* and *Katharina* and describe gene arrangement differences which are particularly useful for each of these approaches, based on: (1) the phylogenetic information inherent in shared gene arrangements; (2) the availability of well-conserved probe sequences for Southern hybridization; and (3) the likelihood of sequences suitable for PCR primers in adjacent genes.

Key words: Mitochondrial DNA, Mollusca, *Mytilus*, *Katharina*, phylogeny, gene order.

INTRODUCTION

Traditional Approaches to Phylogeny

Multicellular animals are grouped with fair confidence into various phyla based on shared general body plans. Establishing evolutionary relationships among the various phyla, and among the various classes and orders within each phylum, is much more speculative. Animal phylogeny is most ambiguous at these high levels, where the fossil record is least complete, homology of morphological structures is least discernible, and long periods of time have erased traces of relatedness.

Paleontological studies are limited in resolving these higher level relationships because of the scarcity of fossils from the very early history of life. By the time of the earliest fossil-rich period, the Cambrian, animal diversity was considerable, with nearly all currently-recognized higher taxa represented. Many early animal forms exhibit unique morphologies not recognizable as intermediate between other groups of animals.

Analysis of morphological data is confounded by the great length of time that has elapsed since these taxa diverged. Determining homologous structures among animals with radically different body plans is contentious. Convergence of morphological structures among various animals is common and difficult to recognize. Hypotheses based on considerations of functional morphology or recapitulation of embryological structures lack the desired rigor and falsifiability. Many alternative, often contradictory models of metazoan evolution have been proposed, all based on interpretations of the same embryological, paleontological, and morphological data sets.

[1] Present address: Department of Cell Biology and Neuroanatomy, University of Minnesota, 4–135 Jackson Hall, 321 Church St. SE, Minneapolis, MN 55455, USA

Nowhere are these difficulties more acute than in determining the relationships among the various classes and orders of the phylum Mollusca, or in determining the sister taxon to this phylum. Although most classification schemes agree on uniting the classes Scaphopoda, Bivalvia, Cephalopoda, Gastropoda, and Monoplacophora into a monophyletic Conchifera, the relationships among these classes are equivocal. Monoplacophora has been suggested as the sister group to Gastropoda based on studies of comparative morphology (Götting, 1980) and paleontological material (Knight & Yochelson, 1960). Salvini-Plawen (1985) views Monoplacophora as basal to a clade of the remaining four conchiferan classes. It has also been suggested that Monoplacophora unites with Gastropoda and Cephalopoda in the trichotomous Cyrtosoma (Runnegar & Pojeta, 1974). Cephalopoda is suggested as the basal conchiferan (Götting, 1980) or as the sister group to Gastropoda (Wingstrand, 1985; Runnegar & Pojeta, 1974). Finally, Scaphopods may be primitive mollusks (Lindberg, 1985) or they may be included in a clade with the Bivalvia and the extinct rostroconchs, the Diasoma (Runnegar & Pojeta, 1974).

Relationships among the non-conchiferans are even more ambiguous. Early classification schemes united Aplacophora with Polyplacophora in the Amphineura, first considering neither to be mollusks (von Ihering, 1876), and later recognizing their molluscan affinity (Pelseneer, 1899). Some modern day classifications view a monophyletic Amphineura as the basal group of mollusks (Pojeta, 1980; Haas, 1981) although some continue to question their inclusion in Mollusca (Fretter & Graham, 1962). Another view places Aplacophora as the primitive, basal class of mollusk, creating a clade of the Polyplacophora and the Conchifera (Götting, 1980; Wingstrand, 1985; Scheltema, 1988). Salvini-Plawen (1985) not only views the Aplacophora in this basal position, but splits Aplacophora into two paraphyletic groups, the Caudofoveata and the Solenogastres, a scheme supported by several later studies (Meglitsch & Schram, 1991; Pearse *et al.*, 1987; Brusca & Brusca, 1990; Barnes, 1987; Nielsen, 1985, 1987). Still other analyses view the Conchifera as the primitive group of mollusks, with the Polyplacophora, Caudofoveata, and Solenogastres secondarily derived (Marcus, 1958; Hadzi, 1953, 1963).

The closest sister taxon to the phylum Mollusca has been suggested variously to be Annelida (Götting, 1980; Vagvolgyi, 1967), Arthropoda (Lemche, 1959a,b; Fretter & Graham, 1962), Sipunculida (Inglis, 1985), Turbellaria (Graham, 1955; Salvini-Plawen, 1972, 1980; Haas, 1981), Echiurida, or Nemertina (see discussion in Salvini-Plawen (1985) and Vagvolgyi (1967)). A clade of consisting of Annelida and Mollusca is suggested to have been derived from flatworms (Hammarsten & Runnström, 1925; Boettger, 1959) or coelenterates (Beklemishev, 1963).

The debates on the relationships among molluscan classes and on the evolutionary origin of Mollusca center on alternative models of morphological change. These models differ in interpreting which types of change are feasible or common, to what extent convergence and parallelism occur, and which structures are homologous among the various metazoan bauplane. Resolution of molluscan relationships would permit many conclusions regarding the evolution of the coelom, serially repeated structures, and body segmentation.

Molecular Phylogenies

Molecular phylogenies based on comparing the sequences of nucleotides or amino acids have offered many new insights into the evolutionary relationships among organisms. During the development of the technology to make this feasible, some hoped that all questions of phylogeny would eventually be answered with a high confidence using these techniques. Although it would be hard to overstate the contribution of molecular approaches to evolutionary biology in recent decades, many questions of organismal relationships remain recalcitrant.

Numerous molecular studies of metazoan relationships have compared the nucleotide sequence of 18S ribosomal RNAs (rRNAs). This molecule was chosen for sequence comparisons in part because its high copy number in the cells of many organisms allows sequence determination directly (without the need for cloning the gene itself) using the enzyme reverse transcriptase. Although they effectively outline the broad pattern of relationships among kingdoms of organisms (Pace *et al.*, 1986; Woese & Fox, 1977; Woese, 1987; Lake, 1988; Sogin, 1991; Wainright *et al.*, 1993), when applied to relationships among metazoans these comparisons have yielded phylogenies that are mutually contradictory and difficult to reconcile with patterns of morphological divergence (Field *et al.*, 1988; Lake, 1990; Patterson, 1989). While only three molluscan classes are represented (Bivalvia, Polyplacophora, and Gastropoda), various methods of phylogenetic analysis of the 18S rRNA data yield contradictory results of the relationships among these classes and their relationships to other animals (see figure 1).

There are numerous limitations of DNA sequence comparisons for phylogeny, especially when considering ancient divergences: (1) With only four possible character states at each aligned position, homoplasy is common and difficult to recognize. (2) These characters states are difficult to reliably polarize as primitive versus derived. (3) Unequal rates of nucleotide substitution in various lineages can lead to erroneous linkages when using distance-based phylogenetic methods, or to the "long-branch attraction" problem when using parsimony-based analyses (Felsenstein, 1978). By counting the number of lineage-specific nucleotide substitutions for 20 taxa for a sample of 1000 rRNA nucleotides, Ghiselin (1988) concluded that substitution rate varies over an 18-fold range among the metazoa. (4) The effects of gene conversion in multigene families may produce patterns of change among groups of organisms that are difficult to deduce (Arnheim, 1983; Sasaki *et al.*, 1987; Walsh, 1985; Coen, Strachan & Dover, 1982). (5) Small subunit rRNAs vary in length by almost a factor of two, so many gaps must be introduced for nucleotide alignment. This

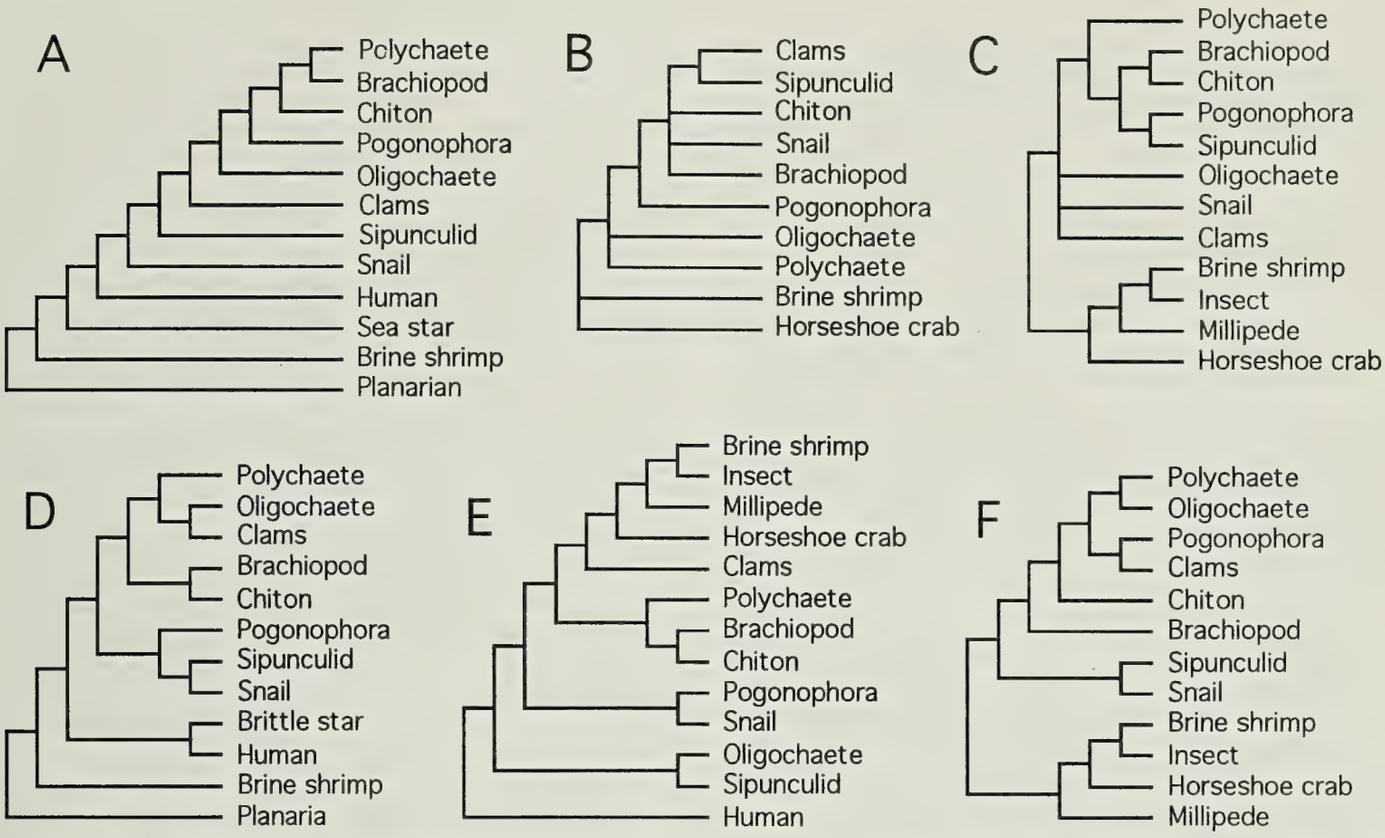


Figure 1. Six evolutionary hypotheses resulting from 18S rRNA comparisons. In some cases, taxa that do not group within the depicted clade have been omitted in order to emphasize the placement of the mollusks. (A) Tree from Field *et al.* (1988), derived using a distance method of analysis for 839 aligned positions. Mollusca is paraphyletic with the inclusion of two annelids, a brachiopod, a pogonophoran, and a sipunculid. Gastropoda is primitive; however, no scaphopods, cephalopods, aplacophorans, or monoplacophorans are included in this analysis. (B) A portion of the rooted evolutionary tree from Lake (1990), using the method of evolutionary parsimony with exactly the same data and alignments as in (A). Mollusca is paraphyletic with the inclusion of a sipunculid and a brachiopod. (C)-(F) Various results of parsimony analyses from Patterson (1989). (C) A portion of the tree produced by a strict consensus of the four shortest trees in a parsimony analysis of 544 aligned nucleotides. Data and alignment are the same as in A and B, but with the addition of two prokaryotic 16S rRNA sequences as outgroups. (D) The tree produced using the same data as for (C), but limiting the taxa analyzed to those depicted, and rooting the tree with the planarian sequence. (E) A portion of the tree produced by limiting the analysis of (C) to only those 273 aligned positions which represent unpaired nucleotides in the rRNA secondary structure. (F) A portion of the tree produced by strict consensus of three equally parsimonious trees, generated by limiting the analysis of (C) to only those 271 aligned positions which represent paired nucleotides in the rRNA secondary structure. Taxa used in these analyses: Clams-*Mya arenaria* and *Spisula solidissima*; Polychaete-*Chaetopteris* sp.; Brachiopod-*Lingula reevi*; Chiton-*Cryptochiton stelleri*; Pogonophora-*Riftia pachyptila*; Oligochaete-*Lumbricus* sp.; Sipunculid-*Goldfingia gouldii*; Snail-*Anisodoris nobilis*; Human-*Homo sapiens*; Sea star-*Asterias forbesi*; Brine shrimp-*Artemia salina*; Planarian-*Dugesia tigrina*; Insect-*Drosophila melanogaster*; Millipede-*Spirobolus marginatus*; Horseshoe crab-*Limulus polyphemus*; Brittle star-*Ophiocoma wendtii*.

alignment is often so ambiguous that different portions of the molecule are compared between the various organisms to eliminate regions which could not be satisfactorily aligned (by subjective criteria). A sequence alignment is fundamentally an hypothesis of homology at each of the aligned positions; changing the alignment of nucleotides can generate very different evolutionary trees. (6) There is disagreement about which parts of the rRNA sequence provide reliable data. Some contend that the most reliable phylogenetic trees are generated by using only the subset of nucleotides that are paired in rRNA secondary structure (Smith, 1989), some argue that the most reliable information is in unpaired nucleotides (Wheeler & Honeycutt, 1988; in a study using 5S and 5.8S rRNA sequences), and some are unable to find a significant difference between the two (Vawter & Brown, 1993). Various weighting schemes for these two subsets of nucleotides lead to different hypotheses of relationships and are based on intractable aspects of the evolutionary history of the molecule (Dixon & Hillis, 1993; Patterson, 1989; see Kraus *et al.*, 1992 for discussion about mitochondrial rRNA sequences). (7) Applying different tree-making methods to the same data frequently yields very different results (e.g., see figure 1). These methods differ in their assumptions of the evolutionary process, and we have no reliable method for discerning

which is most realistic. Most insidious is the tendency to judge results most reliable when congruent to previously accepted hypotheses of relationships. Such circularity questions the potential contribution of molecular phylogenies.

Mitochondrial Genomes

We are interested in discovering a better set of characters for determining molluscan (and other metazoan) relationships. We seek a character set that reliably reflects the evolutionary origin of the phylum Mollusca and the relationships among molluscan classes. To be useful for phylogenetic inference, there are several properties that such a character set must possess: (1) Character states should be determinable for all taxa so no states are shared between organisms as "missing." (2) Characters should be demonstrably homologous among the organisms. (3) Character states should be complex enough so that it would be highly unlikely that the same character state could have arisen independently in two or more lineages. Therefore, identical character states would likely be shared by two or more taxa only because of common ancestry. (4) Character states should change at a rate appropriate for the time span being investigated. Too little change limits resolution; too much change obscures relationships.

We are investigating a set of molecular characters that appear to possess the above properties and show promise to be useful for determining ancient divergences: the arrangement of genes in mitochondrial DNA (mtDNA). MtDNA exists as a discrete genome within the cells of all metazoans, generally as a closed circular DNA of about 14–17 kilobases (kb) (Wolstenholme, 1992a,b; Brown, 1985; Wallace, 1982; Attardi, 1988)[1]. Although much larger mtDNAs are occasionally found in a variety of metazoan taxa, including Mollusca (Moritz, Dowling & Brown, 1987; Snyder *et al.*, 1987; LaRoche *et al.*, 1990), in none of these cases is there any evidence for variation in gene content. MtDNA typically contains one or more large non-coding sequences, which can vary significantly in length among organisms. This region in vertebrates (Clayton, 1991, 1992; Montoya *et al.*, 1982; Bogenhagen, Cairns & Yoza, 1985; King & Low, 1987) and insects (Clary & Wolstenholme, 1985a) has been shown to include elements for the control of replication and transcription. Large variation in lengths of mtDNAs has been due to variation in the length of this major non-coding region (Brown, 1985; Harrison, 1989; Carr, Brothers & Wilson, 1987; Harrison, Rand & Wheeler, 1985; Fauron & Wolstenholme, 1976; Solignac, Monnerot & Mounolou, 1986; Wilkinson & Chapman, 1991; Buroker *et al.*, 1990; Monforte, Barrio & Latorre, 1993) or to duplication of some portion of the mitochondrial genome (Moritz & Brown, 1986, 1987; Zevering *et al.*, 1991).

The arrangement of the genes in mtDNA appears to meet the above four criteria for a character set to be used for phylogenetic inference. The gene content of metazoan mtDNA is well conserved. With few exceptions (Wolstenholme *et al.*, 1987; Okimoto *et al.*, 1991, 1992; Hoffmann, Boore & Brown, 1992), the mtDNA of all animals examined contains the same 37 genes: 2 for the small and large rRNAs of the mitochondrial ribosome (s-rRNA and l-rRNA), 22 for transfer RNAs (tRNAs), and 13 for protein subunits of the enzyme complexes of the inner mitochondrial membrane [cytochrome oxidase subunits I-III (CO1–3), NADH dehydrogenase subunits 1–6 and 4L (ND1–6, 4L), cytochrome b (cytb), and ATP synthase subunits 6 and 8 (ATPase 6, 8)]. Mitochondrial genes are certainly homologous among metazoa, not only because of their functional identity and obvious sequence similarity among metazoans, but also because of the high degree of their similarity in these respects to the mitochondrial genes of non-metazoans. Table 1 shows that each of the 15 protein- or rRNA-encoding genes has a homologue in the mtDNA of one or more non-metazoan organisms. It seems clear that these 15 genes were encoded in mtDNA prior to the origin of Metazoa, and that the gene content of the mtDNA of extant organisms was established prior to this radiation.

The arrangement of mitochondrial genes is a very complex character set for phylogenetic inference, because there are a very large number of possible gene arrangements. Assuming complete positional independence of all genes and two transcriptional orientations, these 37 genes could potentially be joined in greater than 2×10^{52} different arrangements; thus, the probability of the same order arising independently in more than one taxon is vanishingly small. Identical gene arrangements would likely be shared only as a result of common ancestry, making homoplasy rare.

Preliminary studies suggest that the rate of change in mitochondrial gene arrangement is appropriate for resolving ancient divergences. Although mtDNA evolves rapidly in sequence (Brown, George & Wilson, 1979), rearrangements in gene order appear rare. All 37 genes are identically arranged in the mitochondrial genomes of several vertebrates, including *Homo* (Anderson *et al.*, 1981), *Bos* (Anderson *et al.*, 1982), *Mus* (Bibb *et al.*, 1981), *Rattus* (Gadaleta *et al.*, 1989), *Balaenoptera* (Arnason, Gullberg & Widegren, 1991), *Phoca* (Arnason & Johnsson, 1992), *Xenopus* (Roe *et al.*, 1985), *Crossostome* (Tzeng *et al.*, 1992), *Cyprinus* (Chang, Huang & Lo, 1994), and *Gadus* (Johansen, Guddal & Johansen, 1990), although minor rearrangements have taken place in marsupial mammals (Pääbo *et al.*, 1991) and in birds (Desjardins & Morais, 1990, Desjardins, Ramirez & Morais, 1990). The gene arrangement of the mtDNA of the cephalochordate *Branchiostoma floridae* (W. Brown & L. Daehler, unpublished data) is very similar to that of vertebrates. Similarly, mitochondrial genomes have un-

[1] An exception is found among some (but not all) cnidarians, where mtDNA is present as one or two linear molecules totaling about 16 kb (Warrior & Gall, 1985; Bridge *et al.*, 1992).

Table 1. The 15 protein- or rRNA-encoding genes typical of metazoan mtDNA, and a listing of non-metazoan taxa in whose mtDNAs homologous genes have been positively identified. Genes are abbreviated as in the text. Additional genes are also present in these non-metazoan mtDNAs. The non-metazoan species are *Paramecium aurelia, Leishmania tarentolae, Trypanosoma brucei, Chlamydamonas reinhardtii, Marchantia polymorpha, Neurospora crassa, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Aspergillus nidulans,* and *Podospora anserina.* Data are from Boer and Gray, 1991; Brown *et al.*, 1985; Clark-Walker, 1989; Cummings and Delmenico, 1988; Cummings *et al.*, 1990a,b; Dewey *et al.*, 1985; de Zamaroczy and Bernardi, 1986; Dyson *et al.*, 1989; Feagin *et al.*, 1988; Gray and Boer, 1988; Ise *et al.*, 1985; Lang *et al.*, 1983; Oda *et al.*, 1992; Pratje *et al.*, 1989; Pritchard *et al.*, 1990; Simpson *et al.*, 1987; Stuart and Feagin, 1992; Wolf and Del Guidice, 1988; Vahrenholz *et al.*, 1985.

	Cytochrome oxidase			ATPase		Cyt	NADH dehydrogenase							rRNA	
	1	2	3	6	8	b	1	2	3	4L	4	5	6	Sml	Lrg
Metazoa	X	X	X	X	X*	X	X	X	X	X	X	X	X	X	X
Paramecium	X	X				X	X	X	X		X	X		X	X
Leishmania	X	X	X	X		X	X				X	X		X	X
Trypanosoma	X	X	X	X		X	X				X	X		X	X
Chlamydamonas	X					X	X	X			X	X	X	X	X
Marchantia	X	X	X	X		X	X	X	X	X	X	X	X	X	X
Neurospora	X	X	X	X	X	X	X	X	X	X		X	X	X	X
Saccharomyces	X	X	X	X	X	X								X	X
S. pombe	X	X	X	X	X	X								X	X
Aspergillus	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Podospora	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

* ATPase 8 appears to be missing from the mitochondrial genomes of *Mytilus* and nematodes (see text).

dergone little gene rearrangement among echinoderm classes, with only a single large inversion separating the arrangements of sea stars and sea urchins (Jacobs *et al.*, 1988; Cantatore *et al.*, 1989; De Giorgi *et al.*, 1991; Himeno *et al.*, 1987; Smith *et al.*, 1989, 1990, 1993). In each of these deuterostome phyla, organisms which have been separated for over 500 million years share very similar mitochondrial gene arrangements.

The traditionally accepted superphylum, Protostomia, includes three major phyla: Arthropoda, Mollusca, and Annelida. *Drosophila* and *Apis* represent the only genera within arthropods for which the complete mtDNA sequences are published (Clary & Wolstenholme, 1985a; De Bruijn, 1983; Garesse, 1988; Crozier & Crozier, 1993). These two mitochondrial genomes are nearly identically arranged, differing only by a few tRNA gene translocations. Partial gene arrangements for numerous other arthropods (including *Artemia, Locusta, Aedes, Daphnia, Homarus,* and *Limulus*) indicate that gene order is highly conserved in this phylum (Batucus *et al.*, 1988; Hsuchen, Kotin & Dubin, 1984; Dubin, Hsuchen & Tillotson, 1986; McCracken, Uhlenbusch & Gellissen, 1987; Uhlenbusch, McCracken & Gellissen, 1987; D. Stanton, L. Daehler & W. Brown, unpublished data).

Representing pseudocoelomate animals, the mtDNA of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*, have nearly identical gene arrangements (Wolstenholme *et al.*, 1987; Okimoto *et al.*, 1992), although that of a third, *Meloidogyne javanica*, has a radically different arrangement (Okimoto *et al.*, 1991). Partial genome organization has been determined for the parasitic flatworm, *Fasciola hepatica* (Garey & Wolstenholme, 1989) and is unique among animals examined to date. Although gene rearrangements appear to be generally rare within a phylum, there are substantial differences among the gene orders of each of these phyla. Furthermore, the radically different arrangements of mitochondrial genes that exist among several of the taxa suggest that selection for any particular gene order is minimal, and argues against this as a factor that might promote convergence of gene order in separate lineages.

Comparisons of mtDNA gene arrangements may also be useful for phylogeny at lower taxonomic levels. There is evidence that the frequency of rearrangements of the mitochondrial tRNA genes is greater than that of the rRNA and protein genes. For example, the gene orders of two diptera, *Aedes* (Dubin, Hsuchen & Tillotson, 1986, Hsuchen, Kotin & Dubin, 1984) and *Drosophila* (Clary & Wolstenholme, 1985a,b), differ in the relative positions of two tRNA genes and by the relative inversion of a third, but are otherwise identical insofar as can be determined from the partial sequence of *Aedes* mtDNA. The partial gene order determined for another insect, *Locusta*, differs from *Drosophila* by one tRNA rearrangement (MacCracken, Uhlenbusch & Gellissen, 1987; Uhlenbusch, MacCracken & Gellessen, 1987). Eight tRNAs must be repositioned to interconvert the mitochondrial gene arrangements of *Apis* (Crozier & Crozier, 1993) and *Drosophila*. The mitochondrial genome arrangement of several marsupials differs from that of placental mammals by translocations within a cluster of five tRNAs (Pääbo *et al.*, 1991). As more gene orders become available, useful information about phylogenetic relationships among even more closely related animal groups (ordinal and subordinal levels) may occasionally be obtained.

COMPARISON OF THE MITOCHONDRIAL GENOME ARRANGEMENTS OF *MYTILUS EDULIS* (BIVALVIA) AND *KATHARINA TUNICATA* (POLYPLACOPHORA)

We have determined the DNA sequence of 13.9 kb of the 17.1 kb mitochondrial genome of the bivalve *Mytilus edulis*, which is sufficient to identify all 37 mitochondrial genes (Hoffmann, Boore & Brown, 1992). The arrangement of genes in *Mytilus* mtDNA is radically different from those that have been found in other metazoans. With few exceptions, the arrangement of mitochondrial genes appears to be very similar or identical in within-phylum comparisons, so it was initially unclear whether the unusual gene arrangements in *Mytilus* mtDNA was typical of mollusks in general or characteristic of a more restricted group of molluscan taxa. To investigate this, and to evaluate further the potential of mitochondrial genome structure as a useful phylogenetic indicator, we determined the complete mtDNA sequence for the polyplacophoran *Katharina tunicata*. The *Katharina* mitochondrial gene arrangement differs substantially from that of *Mytilus* and is much more similar to the mtDNAs of other coelomate animals, including an annelid (Boore & Brown, manuscript in preparation), and representatives of other classes in the phylum Mollusca (W. Brown, T. Collins & L. Daehler, unpublished data).

The mitochondrial genome of *Mytilus edulis*, in particular, contains several unusual features in comparison with others previously characterized. *Mytilus* mtDNA lacks a gene for ATPase 8. This gene is also absent from the mitochondrial genomes of the three nematodes mentioned above, although presumably these absences represent convergent losses in the *Mytilus* and nematode lineages. *Mytilus* mtDNA encodes 23 tRNAs, one more than the typical metazoan mitochondrial complement. The anticodon of the additional tRNA is complementary to the codons for methionine, giving *Mytilus* mtDNA two methionine tRNA genes. One of these genes specifies a tRNA with the anticodon CAT, which is typical of other metazoan mitochondrial $tRNA^{met}$ genes; the other is nearly unique among all genomes in having the anticodon TAT. The arrangement of these 37 genes is highly unusual, with few gene boundaries shared with any other metazoan so far investigated. The reading frames of the ND1, CO1, and CO3 genes vary significantly in length from those of other metazoans, more so than found in any previous comparisons. However, in other respects this genome is typical of metazoan mtDNA. Aside from the supernumerary tRNAmet genes, its gene content is typically metazoan, its tRNA and rRNA genes are small relative to those found in prokaryotic and eukaryotic nuclei, its gene organization is highly compact, and its genetic code appears to be identical to that employed in several other metazoan mitochondrial systems. All genes are encoded by the same DNA strand, as is the case in other (but not all) metazoans.

Radical variation in the arrangement of mitochondrial genes has been demonstrated previously in comparisons among metazoan phyla, most notably between nematodes and the other phyla examined. However, only minor variation in mitochondrial genome arrangement is usually observed in within-phylum comparisons. It was, therefore, surprising to find the mitochondrial gene arrangement of another mollusk, the polyplacophoran *Katharina tunicata*, to be much more similar those of arthropods, chordates, or echinoderms than to that of *Mytilus*.

Katharina mtDNA encodes the 37 genes typical of metazoan mtDNA, including the gene for ATPase 8 which is absent from *Mytilus* mtDNA. There are at least three possible explanations for the loss of the ATPase 8 gene in the lineage leading to *Mytilus* after its separation from the *Katharina* lineage: (1) The normal function of subunit 8 of the ATP synthase complex is subsumed by another protein subunit; (2) The function of ATPase 8 has become dispensable in the metabolism of *Mytilus* mitochondria; or (3) The ATPase 8 gene has been transferred to the nucleus, and its gene product is now imported into the mitochondria. If it could be determined whether other mollusks share the absence of the ATPase 8 gene from mtDNA, such molecular or metabolic changes could be a very complex derived character, robust for phylogenetic analysis.

In addition to the 22 tRNAs typical of metazoan mitochondrial genomes, *Katharina* mtDNA contains two additional sequences that can be folded into structures resembling tRNAs. If actual tRNAs, their anticodons (AAA and AGA) would presumably recognize the codons UUU and UCU as phenylalanine and serine, respectively. They are, therefore, provisionally identified as $tRNA^{phe(UUU)}$ and $tRNA^{ser(UCU)}$ in figure 2. However, there are several reasons to doubt that they actually function as tRNA genes. The anticodons AAA and AGA are unprecedented in metazoan mtDNA. None of the tRNAs encoded in the mtDNAs of *Katharina*, *Mytilus*, or *Drosophila* have an A in the 3rd ("wobble") position. Both of these putative tRNAs have several mismatches within their stems and neither has a T preceding the anticodon, as is found in all other *Katharina* tRNAs. They do, however account for nearly all of the nucleotides in what would otherwise be unassigned sequence, and their predicted secondary structures are no more aberrant than those of many other mitochondrial tRNAs.

Figure 2 shows the mitochondrial gene arrangements of *Katharina tunicata*, *Mytilus edulis*, and *Drosophila yakuba* (the latter determined by Clary & Wolstenholme, 1985a). If we ignore the positions of tRNA genes, only two rearrangements are necessary to interconvert the gene arrangements of *Katharina* and *Drosophila*: a transposition of the CO3-ND3 segment, and an inversion of the ND6-Cytb segment. The genes encoding tRNAs appear to rearrange at a much higher frequency, as has been noted previously, with numerous tRNAs differing in position between *Drosophila* and *Katharina* for both nearest-neighbor genes, namely those for leu(UUR), lys, asp, gly, ser(AGN), glu, ile, gln, met, and trp.

In contrast, there is little in common when comparing

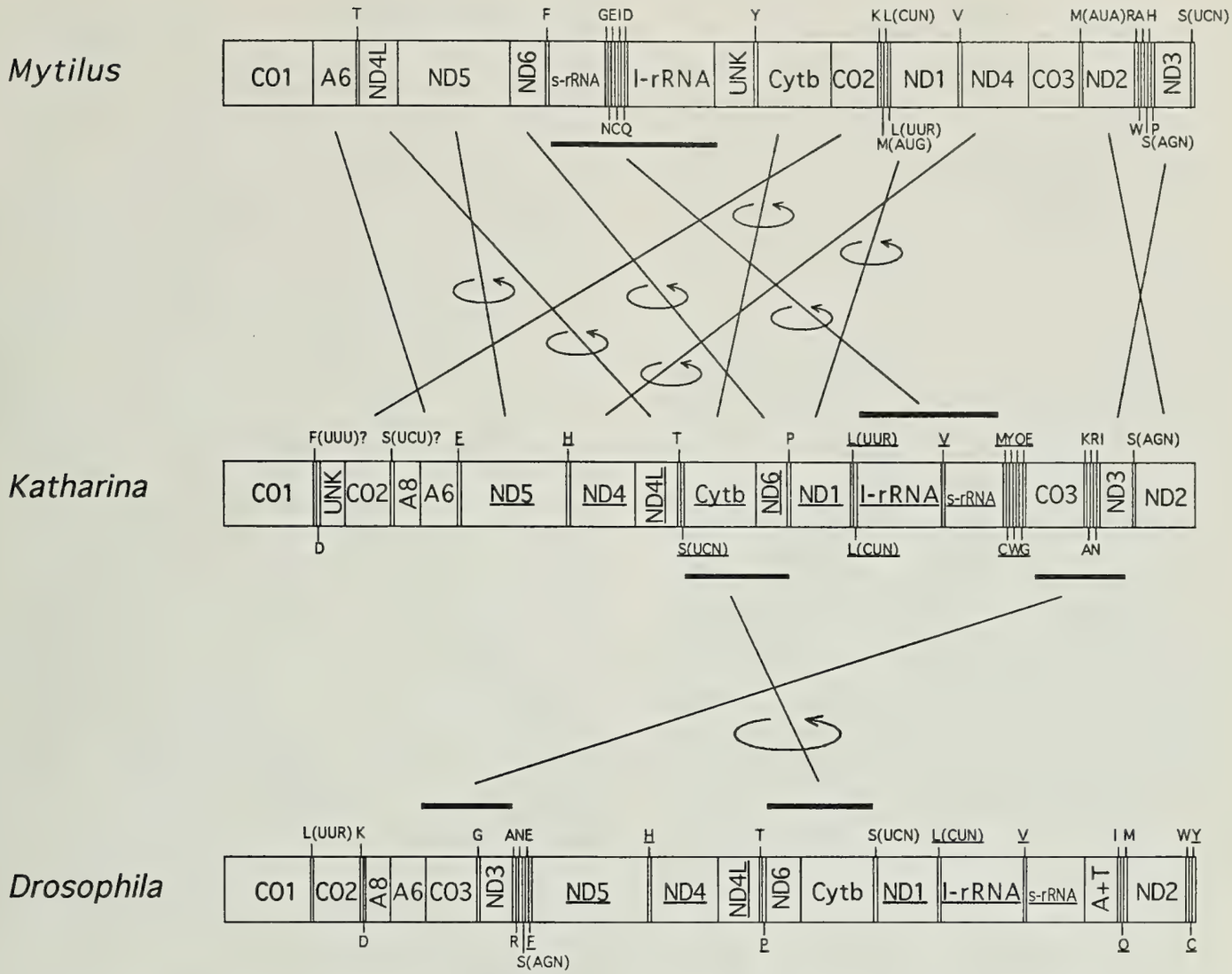


Figure 2. Comparison of mitochondrial gene arrangements among *Katharina tunicata* (Boore & Brown, submitted), *Mytilus edulis* (Hoffmann, Boore & Brown, 1992), and *Drosophila yakuba* (Clary & Wolstenholme, 1985), with each genome aligned starting at the gene for CO1. All genes of *Mytilus* are transcribed from left to right, as are all genes in the *Katharina* and *Drosophila* genomes other than those designated by underlining to signify reverse orientation. Ignoring tRNA position differences, which are numerous, rearrangements are shown by lines connecting gene pairs or blocks of contiguous genes (marked by a bar). Inversions are indicated by a circular arrow. Gene designations are as follows: cytochrome oxidase subunits I-III, CO1–3; NADH dehydrogenase subunits 1–6 and 4L, ND1–6, ND4L; cytochrome b apoenzyme, Cytb; ATP synthase subunits 6 and 8, A6, A8; small and large ribosomal subunit RNAs, s-rRNA, l-rRNA. Transfer RNAs are designated by the one letter code for the corresponding amino acid; the two tRNAs each for serine and leucine are further differentiated by the codon recognized (UCN and AGN for serine; UUR and CUN for leucine). M(AUA) of *Mytilus* mtDNA denotes an additional methionine tRNA and F(UUU)? and S(UCU)? designate additional tRNA-like structures of *Katharina* mtDNA. UNK (unknown) designates the largest unassigned region of the *Mytilus* and *Katharina* mtDNAs. A+T in *Drosophila* designates the A+T rich non-coding region.

the mitochondrial gene arrangement of *Mytilus* with either *Katharina* or *Drosophila*. Ignoring tRNA genes, only the two rRNA genes are in the same order and transcriptional polarity in the three animals. However, in *Drosophila* and *Katharina* mtDNA (and many other metazoans) the two rRNA genes are separated by $tRNA^{val}$, whereas in *Mytilus* mtDNA they are separated by seven tRNAs, none of which is $tRNA^{val}$. The only gene boundaries shared by the *Katharina* and *Mytilus* mitochondrial genomes are those of $tRNA^{thr}$-ND4L and the three gene block, $tRNA^{leu(CUN)}$-$tRNA^{leu(UUR)}$-ND1. The only gene boundaries shared by the *Drosophila* and *Mytilus* mitochondrial genomes are those of CO2-$tRNA^{lys}$, $tRNA^{thr}$-ND4L (although here the relative polarity of $tRNA^{thr}$ is reversed), and $tRNA^{met}$-ND2 (although here the $tRNA^{met}$ of *Mytilus* is has the anticodon TAT whereas the $tRNA^{met}$ of *Drosophila* has the anticodon CAT).

The genes of metazoan mtDNA are typically arranged very compactly. Introns are absent, intergenic nucleotides are few, and genes frequently overlap or end on

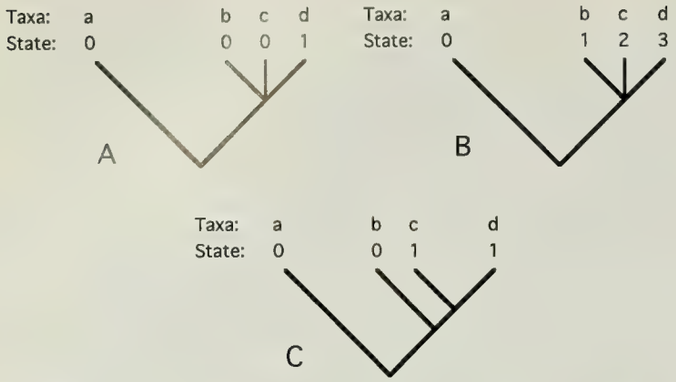


Figure 3. An explanation of the method for analyzing the evolutionary significance of patterns of mitochondrial genome rearrangement. Each taxon is represented by a letter, with "a" designated as the outgroup. Sharing a number for the character state represents sharing a gene boundary; differing numbers indicate that the taxa differ in the gene boundary. Three types of patterns may occur: (A) taxa "b" and "c" share a primitive gene arrangement (symplesiomorphy), since the "0" state existed prior to the origin of "b", "c", or "d". Hence "b" and "c" cannot be united to the exclusion of "d" by this shared gene boundary. (B) taxa "b", "c", and "d" each have unique gene arrangements (autapomorphies). (C) the only pattern of gene arrangements that can be used to unite taxa. Because taxon "b" shares the "0" state with the outgroup, the state is polarized to indicate "c" and "d" share the derived "1" state as a synapomorphy. To place taxon "b" within the clade containing "c" and "d" would be less parsimonious, since it would require either the reversion to the "0" state in taxon "b" or the convergent gain of the "1" state in taxa "c" and "d". Considering the large number of potential character states at each gene boundary, such reversion or convergence to identical states is improbable.

abbreviated stop codons (Wolstenholme, 1992a,b; Brown, 1985; Moritz, Dowling & Brown, 1987; Attardi, 1988). Because genes often abut directly or overlap, any genome rearrangement would require very precise breakage and recombination for the resultant genome to produce functional products. This barrier to recombination has been offered as one possible explanation for the conservation of arrangement of mitochondrial genes over long periods of time (Brown, 1985). In *Mytilus* mtDNA there are five lengthy intergenic sequences, four of which range in size from 79 to 119 nucleotides and the fifth of which is 1.2 kb. It may be that the relatively lengthy regions of DNA without apparent function between genes in *Mytilus* mtDNA enable recombination, accounting for what appears to be an unusually rapid rate of rearrangement, although this remains undemonstrated.

For most metazoans, mtDNA is inherited maternally (Lansman, Avise & Huettel, 1983; Dawid & Blackler, 1972; Gyllensten, Wharton & Wilson, 1985). This is partly due to the vastly greater number of mitochondria in the egg cytoplasm than in the sperm. Sperm cells typically contain only a few mitochondria, and these are localized in the midpiece, which is often excluded from the egg during fertilization. *Mytilus* is a notable exception in this regard. Its mtDNA is frequently inherited biparentally and two or more variant forms of mtDNA often occur within an individual, a condition known as heteroplasmy (Hoeh, Blakely & Brown, 1991; Zouros *et al.*, 1992). Another bivalve, the scallop *Placopecten*, also exhibits frequent heteroplasmy as well as large variations in mitochondrial genome size (Snyder *et al.*, 1987; Gjetvaj, Cook & Zouros, 1992; LaRoche *et al.*, 1990). It is possible that one or more of these unusual features is responsible for the highly derived state of mtDNA in *Mytilus* and, possibly, in other bivalves. This, however, is very speculative, and a much broader survey of bivalve mtDNAs is needed to determine when the radical variation in gene arrangement occurred.

ANALYSIS OF GENE ARRANGEMENTS

It would be overly simplistic and perhaps wrong to suggest that a shared arrangement of mitochondrial genes by itself indicates a close evolutionary relationship. Taxa may share a gene arrangement because they inherited it in an unchanged form which existed ancestral to the divergence of the several taxa being considered (symplesiomorphy; figure 3A); likewise, the gene arrangements of closely related taxa may differ due to a rearrangement that is unique to one of the lineages (autapomorphy; figure 3B). Neither of these patterns of gene arrangement is indicative of phylogeny. The gene arrangements that are useful for phylogenetic inference are those which can be demonstrated, by comparison with appropriate outgroups, to be shared in a derived form (synapomorphy; figure 3C). Admittedly, most phylogenetic branching events will not coincide with mitochondrial gene rearrangements. However, when a derived arrangement is shared by two or more taxa, it is extremely likely to indicate common ancestry.

We cannot infer that *Katharina* and*Mytilus* are distantly related simply because of the great number of differences in the arrangement of their mitochondrial genes. To infer such a distant relationship would be to assume that mtDNA gene rearrangements occur in a clock-like manner, an assumption that existing data refute. The differences in gene order between these two mtDNAs can only be interpreted as an autapomorphy, given the data at hand, and therefore as phylogenetically uninformative. In the same manner, we cannot infer that *Katharina* and *Drosophila* share a more recent common ancestor than *Katharina* and *Mytilus* based on the greater similarity of the *Katharina* and *Drosophila* gene arrangements; this similarity is a symplesiomorphy, and may represent the state of arrangement in the common ancestor of all three taxa. As in figure 3A, all arrangements of the ingroup taxa are equally parsimonious, therefore these characters are not phylogenetically informative.

It is critical that gene arrangements be determined not only for several representatives of each of the classes of Mollusca, but also for all potential outgroup taxa that might be useful for determining whether gene rearrangements are primitive or derived. A larger survey of

animal mitochondrial genomes might reveal intermediate genome arrangements and, perhaps, identify non-molluscan taxa that have these arrangements. However, with increasing numbers of gene arrangements to compare, determining precisely the most parsimonious pattern of rearrangement becomes exponentially more difficult. Techniques are currently being developed to provide computer analysis of genome rearrangements (Sankoff *et al.*, 1990,1992).

COMPARISON OF OTHER ASPECTS OF THE MITOCHONDRIAL GENOMES OF *MYTILUS EDULIS* AND *KATHARINA TUNICATA*

Molecular phylogenies have been limited largely to comparisons of linear sequences of nucleotides or amino acids. An entire field of scientific inquiry has developed from the need to deduce phylogeny most accurately from these sequence comparisons. However, genomes contain many other complex features that can be compared, such as the arrangement of genes, the relative positions of deletions and insertions, the number and position(s) of regulatory sequences, numerous sequence-based aspects of transcription, translation, and DNA replication, variations in the genetic code, and secondary structures of transfer and ribosomal RNAs.

The analysis and comparison of such characteristics among the large and complex nuclear genomes of metazoans will be most informative, but is presently impractical. However, many of these features can be easily accessed for comparison among the much smaller and simpler mitochondrial genomes. Metazoan mtDNA is 25,000 times smaller than the smallest nuclear genome, contains few non-coding nucleotides, has a consistent gene complement and, at least in some organisms, does not appear to undergo genetic recombination (see Wolstenholme, 1992a,b; Brown, 1985; Moritz, Dowling & Brown, 1987; Wallace, 1982).

The tRNA genes are usually interspersed among the protein- and rRNA-coding genes of metazoan mtDNAs. Their product tRNAs fold into complex secondary structures due to internal base-pairing, and it is likely that these structures are present in the polycistronic RNA transcripts and are used as recognition sites by RNA processing enzymes (Ojala *et al.*, 1980; Ojala, Montoya & Attardi, 1981). In *Katharina* mtDNA there are four gene junctions which lack an intervening tRNA, and at each there is a potential secondary structure that positions the start codon of the downstream gene at an identical relative location (Boore & Brown, submitted). If these structures actually form *in vivo*, they may substitute for tRNAs as signals for transcript cleavage. In *Mytilus* mtDNA there are also sequences capable of forming potential secondary structures in the several lengthy intergenic regions, and these may also play a role in the processing of the polycistronic transcript. By investigating mitochondrial RNA processing in mollusks, it may be possible to determine whether these secondary structures actually form, whether they serve as signals for processing enzymes, and whether any of the RNA processing mechanisms are evolutionarily derived for (or within) Mollusca. Such information is, thus, potentially relevant for molluscan phylogeny.

Animal mtDNA uses several variations of the genetic code (see Jukes & Osawa, 1990, 1993, and Wolstenholme, 1992a,b). TGA specifies tryptophan. ATA specifies methionine in all but echinoderm and cnidarian mtDNA. AGA and AGG specify serine in echinoderms, arthropods, nematodes, and platyhelminths, arginine in cnidarians, glycine in ascidians, and are probably stop codons in mammalian mtDNA. AAA usually specifies lysine, but in echinoderm and platyhelminth mtDNA it specifies asparagine. Comparisons of codon usage patterns and protein alignments suggest that both *Mytilus edulis* and *Katharina tunicata* mtDNAs have a genetic code that is identical to that of arthropods. Two major protostome phyla, Mollusca and Arthropoda, are therefore united in this feature.

While nuclear genes initiate translation exclusively with the methionine codon ATG, metazoan mitochondrial genes employ several additional initiation codons, including ATT, ATA, ATC, GTG, TTG, GTT, and ATAA (see Wolstenholme, 1992b). In all but one case, it is unclear whether the initial amino acid of mitochondrial proteins varies with the start codon used, or whether the alternate start codons are somehow recognized by a methionyl-tRNA when they occur as the initial codon of a mRNA. For one human mtDNA gene and transcript, Fearnley and Walker (1987) have determined by sequencing the corresponding protein that ATT in the initiator position specifies methionine, but that it specifies isoleucine when it is in an internal position. *Katharina* mitochondrial genes appear to initiate translation with ATG, ATA, and GTG. *Mytilus* mitochondrial genes appear to initiate translation only with ATG or ATA, although the possible use of other start codons cannot be ruled out due to significant ambiguity in determining the start point of several genes.

Both ATA and ATG code for methionine within the reading frames of mitochondrial proteins. It is not obvious how the single methionyl-tRNA encoded in most mitochondrial genomes discriminates initiation codons, which are translated with N-formyl-methionine, from internal methionine codons, since both may be either ATA or ATG. For this reason, it is intriguing that*Mytilus* mtDNA contains two tRNAs for methionine. However, since the codons expected to pair most efficiently with each of these tRNAs (ATA and ATG; the anticodons of these two tRNAs are UAU and CAU) are present as both initiation codons and in internal positions, the differential use of these tRNAs in initiation and protein extension is not likely.

Termination of translation is also unusual in metazoan mtDNA. In the mtDNA of protostomes all codons are typically used within open reading frames except the stop codons TAA or TAG. However, many genes end with "abbreviated" stop codons of T or TA. In human mitochondria, where it has been investigated, the gene-specific message is precisely cleaved after a T or TA, the first or first and second nucleotides of the terminal codon,

after which the stop codon TAA is completed by polyadenylation of the gene specific message (Ojala, Montoya & Attardi, 1981). Both of the mitochondrial genomes characterized in this work appear to employ this mechanism commonly.

Nuclear encoded tRNAs are invariant for a number of primary and secondary structure features, such as the typical "three-leaf clover" structure, the nucleotides T, pseudo-U, C in one arm (designated the TψC arm) and dihydrouracyl in another (the DHU arm; see Lewin, 1987). The tRNAs of metazoan mtDNA are much more variable, both in primary and secondary structure. Nematode mitochondrial tRNAs are especially unusual; each tRNA is unpaired for the entire TψC arm. (Wolstenholme *et al.*, 1987; Okimoto *et al.*, 1991, 1992). This feature is of great potential use for assessing the hypothesis of monophyly of the Aschelminthes, the group into which nematodes are often placed.

All sequenced metazoan mtDNAs contain two different $tRNA^{ser}$ genes, recognizing codons AGN and UCN, respectively. It is common for one of these, the $tRNA^{ser(AGN)}$, to lack the potential for base-pairing in the DHU arm, and this characteristic is found in both *Katharina* and *Mytilus*. However, in *Katharina* mtDNA the DHU arm is unpaired in the second serine tRNA, $tRNA^{ser(UCN)}$, as well. Since both serine tRNAs must be charged with the same amino acid, perhaps in *Katharina* the DHU portion of the tRNA structure is recognized by the same charging enzyme. If true, and if other molluscan taxa share this shift in tRNA structure and mechanism for serine tRNA charging, this would also be a useful character for phylogenetic analysis. Sequence determination of the serine tRNA genes from other mollusks will allow us to assess this.

Based on analysis of DNA sequence alone, each of the two mollusk classes investigated may encode one or more tRNAs in addition to the normal metazoan complement of 22. *Mytilus* mtDNA may contain an additional tRNA for methionine; *Katharina* mtDNA may contain additional tRNAs for serine and phenylalanine. Further investigation of molluscan mtDNAs will reveal whether the presence of supernumerary tRNA genes is common in this phylum. If it is, then this may also indicate that there are molecular mechanisms in the mitochondrial system of some mollusks that are specific to the evolutionary history of this phylum and that can be used for phylogenetic inference.

Ribosomal RNAs also form elaborate secondary structures through internal base pairing. In comparisons of rRNA gene sequences from various organisms, it is apparent that deletion or addition of large structures in the rRNA has been a common mode of evolution (Clary & Wolstenholme, 1985b; Zwieb, Glotz & Brimacombe, 1981). Perhaps these large scale changes accompany a shift in ribosome functioning. We are currently developing models of secondary structure for the small and large rRNAs of these two mollusk mitochondrial genomes in hopes of identifying structural variations that might be used to infer relationships among molluscan lineages.

FUTURE DIRECTIONS

Initially, it was surprising to find that the arrangement of genes in the mtDNA of *Katharina* was so different from that of *Mytilus*. Although radical variation in mitochondrial gene arrangement has been noted in comparisons among phyla (as in nematode versus coelomate mtDNAs), mitochondrial genome rearrangements appear to be infrequent within phyla. The very different mitochondrial gene arrangements of *Mytilus* and *Katharina* provide numerous character states for investigating molluscan relationships. By screening additional molluscan mitochondrial genomes for gene junctions unique to one of these two arrangements, and by comparing these arrangements with those of non-mollusks, it may be possible to deduce the broad pattern of the evolutionary history of Mollusca. Specifically, the first goal is to investigate mitochondrial gene arrangements in representatives of each of the remaining classes of mollusks. Non-molluscan protostomes must also be investigated to help characterize gene arrangements as primitive or derived.

Investigating gene arrangements by determining complete mtDNA sequences is very costly and laborious. With the knowledge gained from the complete mtDNA sequences of *Katharina* and *Mytilus*, it may be possible to employ less costly and easier methods to screen other molluscan taxa for particular gene arrangements that are likely to be phylogenetically informative. For example, the polymerase chain reaction (PCR) can be used to selectively amplify gene boundaries that are unique to one of these two arrangements, and a large number of animals can be rapidly tested for the arrangement of several of the large, well-conserved genes by Southern hybridization analysis.

The development of DNA amplification via PCR (see Innis *et al.*, 1989) has enabled DNA sequence determination without the difficult and time-consuming procedures of restriction mapping and cloning. As outlined in figure 4, a segment of DNA is amplified to sufficient quantity for gel analysis and DNA sequence determination by employing two oligonucleotides complimentary to flanking sequences. These oligonucleotides serve as primers for the synthesis of new DNA strands by a thermostable DNA polymerase.

The success of a PCR amplification is critically dependent on the complementarity of the oligonucleotide primers to the sequences flanking the DNA to be amplified. The complete mtDNA sequences of *Katharina* and *Mytilus* aid in primer design for screening additional molluscan mtDNAs in two ways. First, since closely related organisms are more likely to share sequence identities, the sequences of the primers can be chosen to match well conserved portions of these two genomes, increasing the likelihood of a successful amplification in the target genome. Second, the gene arrangements of*Mytilus* and *Katharina* mtDNAs give hypotheses of gene arrangement to test on additional animals, since the PCR can only be successful if the primers "face" one

another (see figure 4). Since there is a practical limit to the size of a DNA sequence that can be successfully amplified, primers must be selected in genes that are likely to be closely spaced in molluscan mtDNA. Gene arrangements to be investigated by PCR can be selected based on both the likelihood of finding sequences suitable for PCR primers in adjacent genes and on the amount of phylogenetic information in the sharing of particular gene arrangements.

One strength of this approach is that DNA sequence information is concurrently gained, which can be used in sequence-based phylogenetic analyses as a separate test of relationships. In addition, this technique precisely maps contiguous gene arrangements, and may identify tRNA rearrangements which would be invisible to the technique of Southern hybridization. Crude and highly impure DNA preparations can be used for amplification, including those made from ancient tissues and museum specimens (Thomas, *et al.*, 1990; Kocher *et al.*, 1989). PCR amplification also enables the analysis of DNA from very small organisms and from tiny portions of tissue from rare ones.

The main disadvantage of PCR is that it can not identify novel gene arrangements, but only test for hypothesized arrangements (other than small gene insertions). Primers must be designed to amplify a specific segment, opposing one another over a short segment of DNA. If the flanking sequences to which the primers are designed have rearranged significantly, no amplification will occur. This negative result provides no information, because amplification may fail for numerous reasons in addition to gene rearrangement (e.g., because there have been mutations in a few nucleotides in the region complementary to the primer sequences).

Figure 5 illustrates a gene arrangement that may be amenable to investigation in other molluscan mtDNAs through PCR amplification. The region to be amplified is flanked by the genes for the l-rRNA and cytb. Each of these two genes individually has been successfully amplified by PCR from a variety of organisms (Kocher *et al.*, 1989; D. Stanton and W. Brown, unpublished data). The l-rRNA and cytb genes are well-conserved, based on comparisons among widely divergent taxa; this maximizes the likelihood of finding primer sequences that are useful over a broad taxonomic range.

Determining the arrangement of these particular genes in other mollusks may yield phylogenetically useful information. In *Katharina* and *Drosophila* mtDNAs, ND1 is between l-rRNA and cytb. This represents a symplesiomorphy, since vertebrate mtDNAs share this characteristic as well. The block of genes $tRNA^{L(CUN)}$-$tRNA^{L(UUR)}$-ND1 is shared between the mtDNAs of *Katharina* and *Mytilus*, but in the latter this entire block is translocated to another region of the genome. Any mollusks that share this translocation with *Mytilus* mtDNA are likely to have a common evolutionary history with Bivalvia. In comparing the mtDNAs of *Katharina* and *Drosophila*, the block $tRNA^{Ser(UCN)}$-cytb-ND6-$tRNA^{Pro}$ has been inverted. Any mollusks that share this

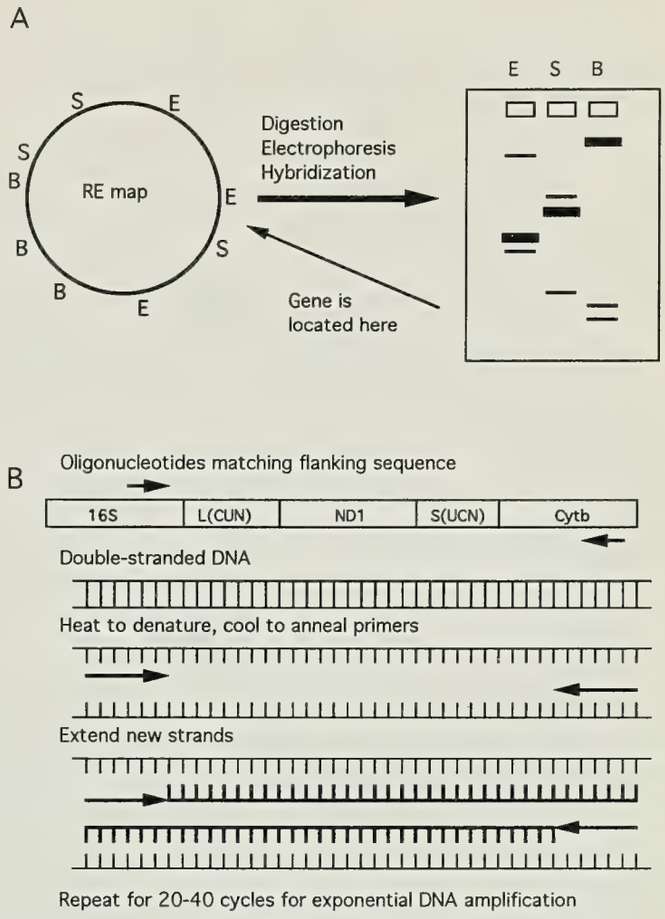


Figure 4. Two alternative techniques for determining the arrangement of mitochondrial genes. (A) The technique of Southern hybridization (Southern, 1975). A cleavage map of the relative locations of restriction enzyme sites in the mtDNA is constructed. Each enzyme recognizes a particular short sequence of DNA (4–8 bp). E, S, and B in this figure represent the locations of three independent restriction enzyme (EcoRI, SalI, and BamHI) cleavage sites on the circular map of the mitochondrial genome. The mtDNA is cleaved with each enzyme, the fragments produced are separated according to size by electrophoresis through an agarose gel, the fragments in the gel are transferred to a membrane and probed using a radiolabeled DNA fragment that contains all or part of the gene of interest. The fragment patterns generated by each restriction enzyme are labeled E, S, and B on the depiction of the gel. The radiolabeled probe will hybridize only to the fragments of the mtDNA that contain the corresponding gene, shown in bold. This information, when correlated with the mtDNA cleavage map, provides the gene's position in the mitochondrial genome. (B) The technique of PCR amplification (Innis *et al.*, 1987). Oligonucleotide primers that are complementary to the DNA sequence flanking the region of interest are determined and synthesized. The double-stranded template DNA is heat denatured, mixed with a vast excess of the oligonucleotides, then cooled to allow annealing of the oligonucleotides to the template. The oligonucleotides serve as primers for the synthesis of new strands of DNA in a reaction using thermostable DNA polymerase. This cycle of heat (denaturation) and cool (anneal and synthesize) is repeated many times (typically 20–40) to exponentially amplify the DNA region between the primers to provide amounts that allow manipulation and determination of the DNA sequence.

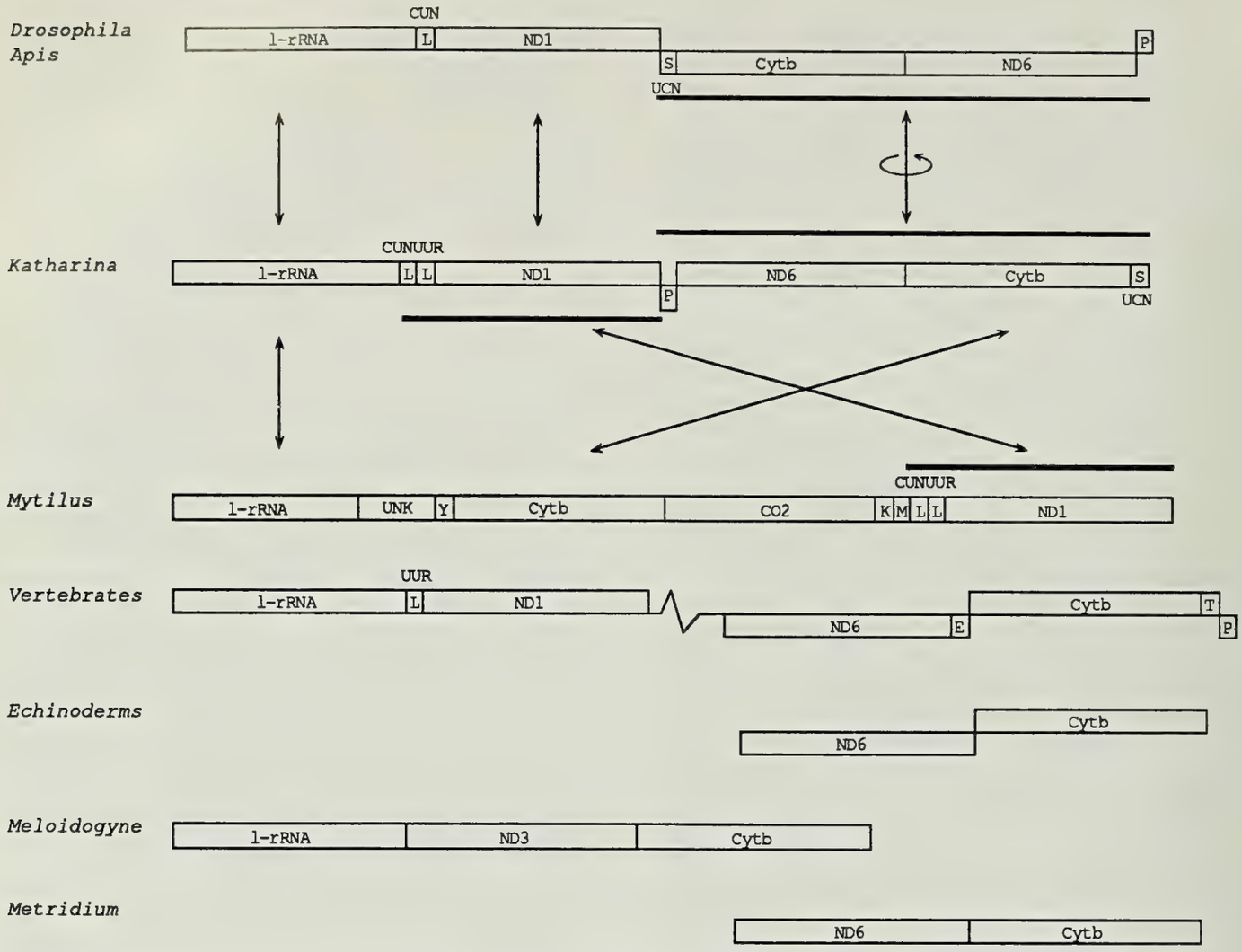


Figure 5. The arrangements of several mitochondrial genes particularly amenable to investigation by PCR (see text) and potentially informative for molluscan phylogeny. Gene abbreviations are as in figure 2. Genes are transcribed from left to right except those depicted below the main line to designate opposite orientation. The broken line shown for vertebrate mtDNA indicates a large, undepicted portion of the genome. The positional relationships of genes in the insect and mollusk mtDNAs are depicted as in figure 2; the arrangement of these genes in mtDNAs of other organisms are shown for comparison (references in text except for *Metridium*, D. R. Wolstenholme, personal communication). Genes are not drawn to scale.

inversion with *Katharina* are likely to also share a common ancestor with Polyplacophora. If the arrangements of these genes is determined for representatives of all molluscan classes, it should be possible to formulate a phylogenetic hypothesis for the Mollusca, and to test the hypothesis with additional gene arrangement data.

The second technique for rapidly screening mitochondrial gene arrangements is Southern hybridization (Southern, 1975), which localizes genes relative to a physical map of the mtDNA. This technique is outlined and described in figure 4. Cleavage sites for restriction enzymes which recognize specific short sequences of DNA determine the physical map. Restriction endonuclease cleavage generates DNA fragments, which are separated by size using gel electrophoresis, visualized, and related back to their position in the mtDNA. Gene-specific probes that are labeled with a radioisotope are exposed to the gel-separated DNA bands under hybridizing conditions. The probes hybridize specifically to the DNA bands which include the probe gene. The position of this gene can then be correlated to the physical map of the mtDNA and localized relative to other probed genes. The subcloning necessary for determining the complete mtDNA sequence of *Katharina* and *Mytilus* provides the gene-specific probes necessary to this technique.

One limitation of the Southern hybridization technique is the requirement for a detailed cleavage map of the mtDNA. This is most effectively accomplished if the mtDNA can be recovered in pure form and in large quantity, which is often difficult and may be impossible for some small or rare organisms. The resolution of the gene map will be limited by the spacing of the restriction enzyme cleavage sites. The main advantages of Southern hybridization are that 1) the relative location of widely spaced genes can be determined, whereas PCR can only be applied to contiguous blocks of closely adjacent genes

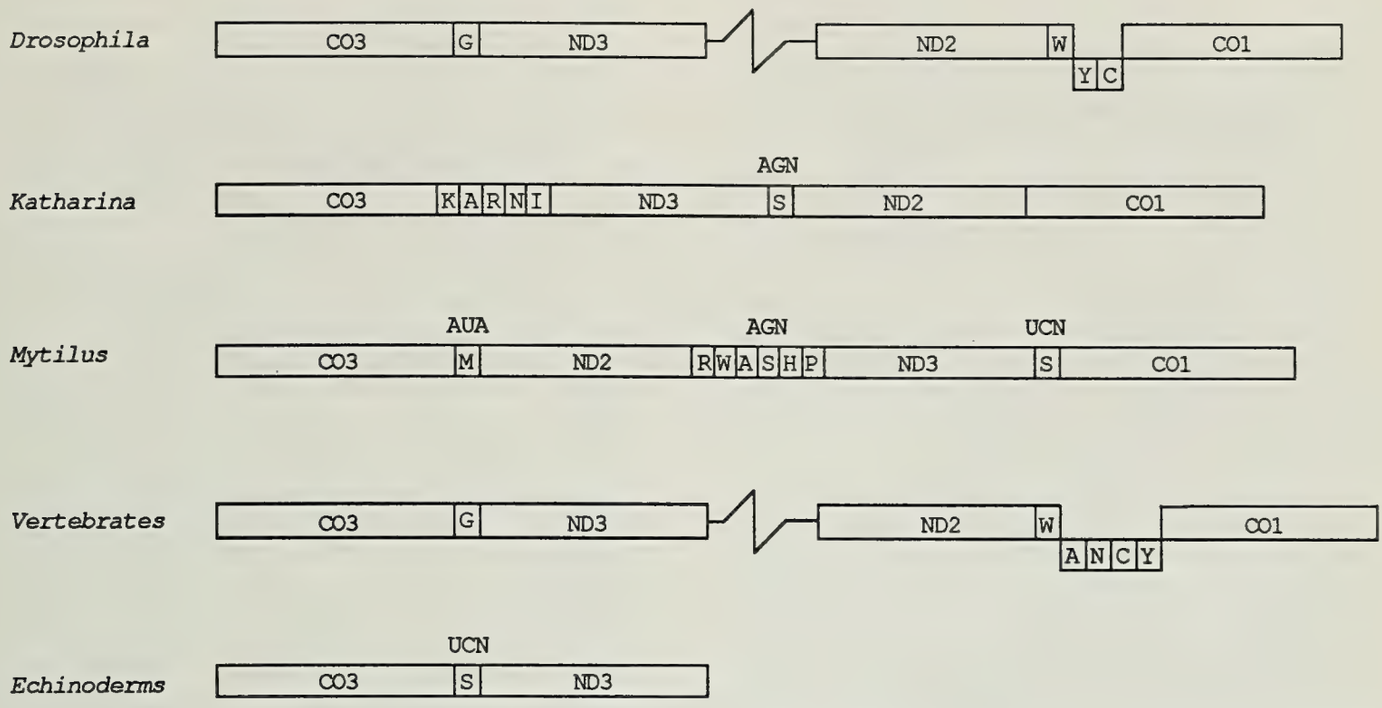


Figure 6. The arrangements of several mitochondrial genes promising for Southern hybridization analysis (see text). Gene abbreviations are as in figure 2; genes are transcribed from left to right except those depicted below the main line to designate opposite orientation (references in text). The broken lines shown for *Drosophila* and vertebrate mtDNAs indicates a large, undepicted portion of the genome.

and 2) no prior hypothesis of gene arrangement is necessary for Southern hybridization, in contrast to the case for PCR.

Figure 6 shows the relative arrangement of four genes in the *Mytilus*, *Katharina*, and *Drosophila* mitochondrial genomes that may be especially useful for investigating molluscan relationships. In general, the gene arrangements of *Katharina* and *Mytilus* are difficult to relate. However, discounting tRNA genes, the arrangement of CO3, ND2, ND3, and CO1 differ only in that ND2 and ND3 have exchanged positions. The arrangement of these genes in *Drosophila* mtDNA is similar in that CO3 is near ND3 and ND2 is near CO1, but these two pairs of genes are separated by approximately 10 kb of DNA sequence. Each of these four genes is well conserved enough to expect that they could be detected in the mtDNA of other mollusks by hybridization to probes of *Mytilus* or *Katharina* mtDNA. Determining the arrangement of these genes in the mtDNAs of other mollusks might suggest whether the rearrangement that brought ND3 near to ND2 occurred near the base of the molluscan radiation. If so, it would provide a synapomorphy suggesting the monophyly of Mollusca. It would also inform us about whether other classes of mollusks share with *Mytilus* the derived condition of inverting the positions of ND2 and ND3, thus signaling a common ancestry with Bivalvia.

Admittedly, most evolutionary relationships will not be resolved by comparisons of mitochondrial gene arrangements. Genome rearrangements may not have occurred during the period of shared history, or subsequent rearrangements may have erased similarity. The main advantage of this data set is that relationships are very reliably inferred when accompanied by a genome rearrangement. The complete sequences of the mtDNAs of *Katharina* and *Mytilus* facilitate even more rapid investigation of the patterns of mitochondrial genome rearrangments among mollusks. As more mtDNAs are investigated for gene arrangement and other complex molecular characteristics, higher level relationships among mollusks and among other taxa may be resolved.

LITERATURE CITED

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. De Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden, and I. G. Young. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457–465.

Anderson, S., M. H. L. De Bruijn, A. R. Coulson, I. C. Eperon, F. Sanger and I. G. Young. 1982. Complete sequence of bovine mitochondrial dna: conserved features of the mammalian mitochondrial genome. Journal of Molecular Biology 156:683–717.

Arnason, U., A. Gullberg and B. Widegren. 1991. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. Journal of Molecular Evolution 33(6):556–568.

Arnason, U. and E. Johnsson. 1992. The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina*. Journal of Molecular Evolution 34:493–505.

Arnheim, N. 1983. Concerted evolution of multigene families. *In*: M. Nei and R. Keohn (eds.). Evolution of Genes and

Proteins. Sinauer Associates, Inc., Sunderland, Massachusetts, pp.38–61.

Attardi, G. 1988. Biogenesis of mitochondria. Annual Review of Cell Biology 4:289–333.

Barnes, R. D. 1987. Invertebrate Zoology, 5th edition. W.B. Saunders, Philadelphia, 893 pp.

Batucas, B., R. Garresse, M. Calleja, J. R. Valverde and R. Marco. 1988. Genome organization of *Artemia* mitochondrial DNA. Nucleic Acids Research 16:6515–6529.

Bibb, M. J., R. A. Van Etten, C. T. Wright, M. W. Walberg and D. A. Clayton. 1981. Sequence and gene organization of mouse mitochondrial DNA. Cell 26:167–180.

Beklemishev, V. N. 1963. On the relationship of the *Turbellaria* to other groups of the animal kingdom. *In*: E. C. Dougherty (ed.). The Lower Metazoa. University of California Press, Berkeley, pp. 234–244.

Boer, P. and M. Gray. 1991. Short dispersed repeats localized in spacer regions of *Chlamydamonas reinhardtii* mitochondrial DNA. Current Genetics 19:309–312.

Boettger, C. R. 1959. Comments to H. Lemche: Protostomian interrelationships in the light of *Neopilina*. Proceedings of the XVth International Congress of Zoology, London, pp.386–389.

Bogenhagen, D. F., S. S. Cairns and B. K. Yoza. 1985. Nucleotide sequences involved in the control of transcription and displacement loop DNA synthesis for *Xenopus laevis* mtDNA. *In*: E. Quagliariello *et al.* (eds.). Achievements and Perspectives of Mitochondrial Research, Vol. II: Biogenesis. Elsevier Science Publishers, pp. 175–182

Boore, J. L. and W. M. Brown. Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata*, submitted.

Boore, J. L. and W. M. Brown. Complete DNA sequence of the mitochondrial genome of the annelid, *Lumbricus terrestris*, manuscript in preparation.

Bridge, D., C. W. Cunningham, B. Schierwater, R. Desalle and L. W. Buss. 1992. Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. Proceedings of the National Academy of Sciences, USA 89:8750–8753.

Brown, T., R. Waring, C. Scazzocchio and R. Davies. 1985. The *Aspergillus nidulans* mitochondrial genome. Current Genetics 9:113–117.

Brown, W. M. 1985. The mitochondrial genome of animals. *In*: R. J. MacIntyre (ed.). Molecular Evolutionary Genetics. Plenum Press, New York, pp. 95–130.

Brown, W. M., M. George and A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences, USA 76:1967.

Brusca, R. C. and G. J. Brusca. 1990. Invertebrates. Sinauer, Sunderland, Massachusetts, 922 pp.

Buroker, N., J. Brown, T. Gilbert, P. O'Hara, A. Beckenbach, W. Thomas and M. Smith. 1990. Length heteroplasmy of sturgeon mitochondrial DNA: An illegitimate elongation model. Genetics 124:157–163.

Cantatore, P., M. Roberti, G. Rainaldi, M. N. Gadaleta and C. Saconne. 1989. The complete nucleotide sequence, gene order and genetic code of the mitochondrial genome of *Paracentrotus lividus*. Journal of Biological Chemistry 264: 10965–10975.

Carr, S. M., A. J. Brothers and A. C. Wilson. 1987. Evolutionary inferences from restriction maps of mitochondrial DNA from nine taxa of *Xenopus* frogs. Evolution 41:176.

Chang, Y.-s , F.-l. Huang and T.-b. Lo. 1994. The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. Journal of Molecular Evolution 38:138–155.

Clark-Walker, G. 1989. *In vivo* rearrangement of mitochondrial DNA in *Saccharomyces cerevisiae*. Proceedings of the National Academy of Sciences, USA 86:8847–8851.

Clary, D. O. and D. R. Wolstenholme. 1985a. The mitochondrial DNA molecule of *Drosophila yakuba*: Nucleotide sequence, gene organization, and genetic code. Journal of Molecular Evolution 22:252–271.

Clary, D. O. and D. R. Wolstenholme. 1985b. The ribosomal RNA genes of *Drosophila* mitochondrial DNA. Nucleic Acids Research 13(11):4029–4044.

Clayton, D. A. 1991. Replication and transcription of vertebrate mitochondrial DNA. Annual Review of Cell Biology 7:453–478.

Clayton, D. A. 1992. Transcription and replication of animal mitochondrial DNA. *In*: K. W. Jeon and D. R. Wolstenholme (eds.). Mitochondrial Genomes. International Review of Cytology 141:217–232.

Coen, E., T. Strachan and G. Dover. 1982. Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. Journal of Molecular Biology 158:17–35.

Crozier, R. H and Y. C. Crozier. 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. Genetics 133:97–117.

Cummings, D. and J. Domenico. 1988. Sequence analysis of mitochondrial DNA from *Podospora anserina*. Journal of Molecular Biology 204:815–839.

Cummings D., F. Michel, J. Domenico and K. McNally. 1990a. DNA sequence analysis of the mitochondrial ND4L-ND5 gene complex from *Podospora anserina*: Duplication of the ND4L gene within its intron. Journal of Molecular Biology 212(2):269–86.

Cummings, D., K. McNally, J. Domenico and E. Matsuura. 1990b. The complete DNA sequence of the mitochondrial genome of *Podospora anserina*. Current Genetics 17:375–402.

Dawid, I. B. and A. W. Blackler. 1972. Maternal and cytoplasmic inheritance of mitochondrial DNA in *Xenopus*. Developmental Biology 29:152–161.

De Bruijn, M. H. L. 1983. *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. Nature 304:234–241.

De Giorgi, C., C. Lanave, M. D. Musci and C. Saccone. 1991. Mitochondrial DNA in the sea urchin *Arbacia lixula*: evolutionary inferences from nucleotide sequence analysis. Molecular Biology and Evolution 8:515–529.

de Zamaroczy, M and G. Bernardi. 1986. The primary structure of the mitochondrial genome of *Saccharomyces cerevisiae*–a review. Gene 47:155–177.

Desjardins, P., and R. Morais. 1990. Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. Journal of Molecular Biology 212:599–634.

Desjardins, P., V. Ramirez and R. Morais. 1990. Gene organization of the Peking duck mitochondrial genome. Current Genetetics 17:515–518.

Dewey, R., C. Leving and D. Timothy. 1985. Nucleotide sequence of ATPase subunit 6 gene of maize mitochondria. Plant Physiology 79:914–919.

Dixon, M. T. and D. M. Hillis. 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. Molecular Biology and Evolution 10(1):256–267.

Dubin, D. T., C. -C. Hsuchen and L. E. Tillotson. 1986. Mosquito mitochondrial tRNAs for valine, glycine and glutamate: RNA and gene sequences and vacinal genome organization. Current Genetics 10:701–707.

Dyson, N., T. Brown, J. Ray, R. Waring, C. Scazzocchio and R. Davies. 1989. Processing of mitochondrial RNA in *Aspergillus nidulans*. Journal of Molecular Biology 208: 587–599.

Fauron, C. M. -R. and D. R. Wolstenholme. 1976. Structural heterogeneity of mitochondrial DNA molecules within the genus *Drosophila*. Proceedings of the National Academy of Sciences, USA 73:3623.

Feagin, J., J. Abraham and K. Stuart. 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. Cell 53:413–422.

Fearnley I. and J. Walker. 1987. Initiation codons in mammalian mitochondria: differences in genetic code in the organelle. Biochemistry 26(25):8247–8251.

Felsenstein, J. 1978. Cases in which parsimony and compatability methods may be positively misleading. Systematic Zoology 27:401–410.

Field, K. G., G. J. Olsen, D. J. Lane, S. J. Giovannoni, M. T. Ghiselen, E. C. Raff, N. R. Pace and R. A. Raff. 1988. Molecular phylogeny of the animal kingdom. Science 239: 748–753.

Fretter, V. and A. Graham. 1962. British prosobranch molluscs. The Ray Society, London, 755 pp.

Gadaleta, G., G. Pepe, G. De Candia, C. Quagliariello, E. Sbis and C. Saccone. 1989. The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. Journal of Molecular Evolution 28:497–516.

Garesse, R. 1988. *Drosophila melanogaster* mitochondrial DNA: Gene organization and evolutionary considerations. Genetics 118:649–663.

Garey, J. R. and D. R. Wolstenholme. 1989. Platyhelminth mitochondrial DNA: Evidence for early evolutionary origin of a $tRNA^{ser(AGN)}$ that contains a dihydrouridine arm replacement loop and of serine-specifying AGA and AGG codons. Journal of Molecular Evolution 28:374–387.

Ghiselin, M. T. 1988. The origin of molluscs in the light of molecular evidence. Oxford Surveys in Evolutionary Biology 5:66–95.

Gjetvaj, B., D. I. Cook and E. Zouros. 1992. Repeated sequences and large scale size variation of mitochondrial DNA: A common feature among scallops (Bivalvia: Pectinidae). Molecular Biology and Evolution 9:106–124.

Gütting, K. -J. 1980. Origin and relationships of the Mollusca. Z. zool. Syst. Evolut. -forsch 18:24–27.

Graham, A. 1955. Molluscan diets. Proceedings of the Malacological Society of London 31:144–159.

Gray, M. and P. Boer. 1988. Organization and expression of algal (*Chlmydamonas reinhardtii*) mitochondrial DNA. Philosophical Transactions of the Royal Society of London 319:135–147.

Gyllensten, U., D. Wharton and A. C. Wilson. 1985. Maternal inheritance of mitochondrial DNA during backcrossing of two species of mice. Journal of Heredity 76:321.

Haas, W. 1981. Evolution of calcareous hardparts in primitive molluscs. Malacologia 21:403–418.

Hadzi, J. 1953. An attempt to reconstruct the system of animal classification. Systematic Zoology 2:145–154.

Hadzi, J. 1963. The evolution of the Metazoa. Macmillan, New York, 499 pp.

Hammarsten, O. and J. Runnström. 1925. Zur Embryologie von *Acanthochiton discrepans* Brown. Zool. Jahrb., Abt. Anat. Ontog. Tiere 47:261–318.

Harrison, R. G., D. M. Rand and W. C. Wheeler. 1985. Mitochondrial DNA size variation within individual crickets. Science 228:1446–1448.

Harrison, R. G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends in Ecology and Evolution 4:6–11.

Himeno, H., H. Masiki, T. Kawai, T. Ohta, I. Kumagai, K. Miura and K. Waranabe. 1987. Unusual genetic codes and a novel gene structure for $tRNA^{ser(AGY)}$ in starfish mitochondrial DNA. Gene 56:219–230.

Hoeh, W. R., K. H. Blakley and W. M. Brown. 1991. Heteroplasmy suggests limited biparental inheritance of *Mytilus* mitochondrial DNA. Science 251:1488–1490.

Hoffman, R. J., J. L. Boore and W. M. Brown. 1992. A novel mitochondrial genome organization for the blue mussel, *Mytilus* edulis. Genetics 131:397–412.

Hsuchen, C. -C., R. M. Kotin and D. T. Dubin. 1984. Sequences of the coding and flanking regions of the large ribosomal subunit RNA gene of mosquito mitochondria. Nucleic Acids Research 12:7771–7785.

Ihering, H. von. 1876. Versuch eines natürlichen Systemes der Mollusken. Jahrbucher der Deutschen malacozoologischen Gesellschaft 3:97–148.

Inglis, W. G. 1985. Evolutionary waves; patterns in the origins of animal phyla. Australian Journal of Zoology 33:153–178.

Innis, M., D. Gelfand, J. Sninsky and T. White (eds.). 1989. PCR Protocols: A Guide to Methods and Applications. Academic Press, Inc., San Diego, CA, 482 pp.

Ise, W., H. Haiker and H. Weiss. 1985. Mitochondrial translation of subunits of the rotenone-sensitive NADH-ubiquinone reductase in *Neurospora crassa*. European Molecular Biology Organization Journal 4:2075–2080.

Jacobs, H. T., D. J. Elliott, V. B. Math and A. Farquarson. 1988. Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. Journal of Molecular Biology 202:185–217.

Johansen, H., P. H. Guddal and T. Johansen. 1990. Organization of the mitochondrial genome of Atlantic cod, *Gadus morhua*. Nucleic Acids Research 18:411–419.

Jukes, T. H. and S. Osawa. 1990. The genetic code in mitochondria and chloroplasts. Experientia 46(11–12):1117–1126.

Jukes, T. H. and S. Osawa. 1993. Evolutionary changes in the genetic code. Comparative Biochemistry and Physiology 106B(3):489–494.

King, T. C. and R. L. Low. 1987. Mapping of control elements in the displacement loop region of bovine mitochondrial DNA. Journal of Biological Chemistry 13:6204–6213.

Knight, J. B. and E. L. Yochelson. 1960. Monoplacophora. *In*: R. C. Moore (ed.). Treatise on Invertebrate Paleontology, I, Mollusca. University of Kansas, Lawrence, pp. 177–184.

Kocher T., W. K. Thomas, A. Meyer, S. Edwards, S. Pääbo, F. Villablanca and A. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceedings of the National Academy of Sciences, USA 86(16):6196–6200.

Kraus, F., L. Jarecki, M. M. Miyamoto, S. M. Tanhauser and P. J. Laipis. 1992. Mispairing and compensational changes during the evolution of mitochondrial ribosomal RNA. Molecular Biology and Evolution 9(4):770–774.

Lake, J. 1988. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. Nature 331: 184–186.

Lake, J. 1990. Origin of the Metazoa. Proceedings of the National Academy of Sciences, USA 87:763–766.

Lang, B., F. Ahne, S. Distler, H. Trinkl, F. Kaudewitz and K. Wolf. 1983. Sequence of the mitochondrial DNA, arrangements of genes and processing of their transcripts in *Schizosaccharomyces pombe*. *In*: R.J. Schweyen, K. Wolf and F. Kaudewitz (eds.). Mitochondria 1983: Nucleo-Mitochondrial Interactions. de Gruyter, Berlin, pp. 313–329.

Lansman, R. A., J. C. Avise and M. D. Huettel. 1983. Critical experimental test of the possibility of "paternal leakage" of mitochondrial DNA. Proceedings of the National Academy of Sciences, USA 80:1969–1971.

LaRoche, J., M. Snyder, D. I. Cook, K. Fuller and E. Zouros. 1990. Molecular characterization of a repeat element causing large-scale size variation in the mitochondrial DNA of the sea scallop *Placopecten magellanicus*. Molecular Biology and Evolution 7:45–64.

Lemche, H. 1959a. Molluscan phylogeny in the light of *Neopilina*. Proceedings of the XVth International Congress of Zoology, London, pp. 380–381.

Lemche, H. 1959b. Protostomian interrelationships in the light of *Neopilina*. Proceedings of the XVth International Congress of Zoology, London, pp. 381–389.

Lewin, B. 1987. Genes III. Wiley & Sons, Inc., New York, pp. 122–143.

Lindberg, D. R. 1985. Aplacophoran, monoplacophorans, polyplacophorans, and scaphopods: the lesser classes. *In*: D. J. Bottjer, C. S. Hickman and P. D. Ward (eds.). Mollusks, Notes for a Short Course. University of Tennessee Department of Geological Sciences Studies in Geology no. 13, pp. 230–247.

Maccracken, A., I. Uhlenbusch and G. Gellissen. 1987. Structure of the cloned *Locusta migratoria* mitochondrial genome: restriction mapping and sequence of its ND–1 (URF–1) gene. Current Genetics 11:625–630.

Marcus, E. 1958. On the evolution of the animal phyla. Quarterly Review of Biology 33:24–58.

Meglitsch, P. A. and F. R. Schram. 1991. Invertebrate Zoology, 3rd edition. Oxford University Press, New York.

Monforte, A., E. Barrio and A. Latorre. 1993. Characterization of the length polymorphism in the A+T region of the *Drosophila obscura* group species. Journal of Molecular Evolution 36(3):214–223.

Montoya, J., T. Christianson, D. Levens, M. Rabinowitz and G. Attardi. 1982. Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. Proceedings of the National Academy of Sciences, USA 79:7195–7199.

Moritz, C., and W. M. Brown. 1986. Tandem duplications of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science 233:1425–1427.

Moritz, C. and W. M. Brown. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proceedings of the National Academy of Sciences, USA 84:7183–7187.

Moritz, C., T. E. Dowling and W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annual Review of Ecology and Systematics 18:269–292.

Nielsen, C. 1985. Animal phylogeny in the light of the trochaea theory. Biological Journal of the Linnaeus Society 25:243–299.

Nielsen, C. 1987. Structure and function of metazoan ciliary bands and their phylogenetic significance. Acta Zoologica 68:205–262.

Oda, K., K. Yamato, E. Ohta, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, T. Kanegae, Y. Ogura, T. Kohchi and K. Ohyama. 1992. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. Journal of Molecular Biology 223: 1–7.

Ojala, D., C. Merkel, R. Gelfand and G. Attardi. 1980. The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. Cell 22:393–403.

Ojala, D., J. Montoya and G. Attardi. 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–474.

Okimoto, R., H. M. Chamberlin, J. L. Macfarlane and D. R. Wolstenholme. 1991. Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (*Meloidogyne*): nucleotide sequences, genome location and potential for host race identification. Nucleic Acids Research 19:1619–1626.

Okimoto, R., J. L. Macfarlane, D. O. Clary and D. R. Wolstenholme. 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics 130(3):471–498.

Pääbo, S., W. K. Thomas, K. M. Whitfield, Y. Kumazawa and A. C. Wilson. 1991. Rearrangements of mitochondrial transfer RNA genes in marsupials. Journal of Molecular Evolution 33:426–430.

Pace, N. R., G. J. Olsen and C. R. Woese. 1986. Ribosomal RNA phylogeny and the primary lines of evolutionary descent. Cell 45:325–326.

Patterson, C. 1989. Phylogenetic relationships of major groups: conclusions and prospects. *In*: B. Fernholm, K.Bremer and H. Jrnvall (eds.). The Hierarchy of Life. Elsevier Science Publishers, B. V.(Biomedical Division), Amsterdam, pp. 471–488.

Pearse, V., J. Pearse, M. Buchsbaum and R. Buchsbaum. 1987. Living invertebrates. Boxwood Press, Pacific Grove, California.

Pelseneer, P. 1899. Recherches morphologiques et phylogénétiques sur les mollusques archaiques. Mémoires Couronnées et Mémoires Savants trangers Academie Belgique 57:1–113.

Pojeta, J. 1980. Molluscan Phylogeny. Tulane Studies in Geology and Paleontology 16(2):55–80.

Pratje, E., C. Vahrenholz, S. Buhler, and G. Michaelis. 1989. Mitochondrial DNA of *Chlamydomonas reinhardtii*: the ND4 gene encoding a subunit of NADH dehydrogenase. Current Genetics 16(1):61–64.

Pritchard, A., J. Seilhamer, R. Mahalingam, C. Sable S. Venuti and D. Cummings. 1990. Nucleotide sequence of the mitochondrial genome of *Paramecium*. Nucleic Acids Research 18(1):173–180.

Roe, B. A., D. -P. Ma, R. K. Wilson and J. F. -H. Wong. 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. Journal of Biological Chemistry 260:9759–9774.

Runnegar, B. and J. Pojeta. 1974. Molluscan phylogeny: the palaeontological viewpoint. Science 186:311–317.

Salvini-Plawen, L. von. 1972. Zur Morphologie und Phylogenie der Mollusken; die Beziehungen der Caudofoveata und der Solenogastres als Aculifera, als Mollusca, und als Spiralia. Zeitschrift fr wissenschaftliche Zoologie 184:205–394.

Salvini-Plawen, L. von. 1980. A reconsideration of systematics in the Mollusca (phylogeny and higher classification). Malacologia 19:249–278.

Salvini-Plawen, L. von. 1985. Early evolution and the primitive groups. *In* E. R. Trueman and M. R. Clarke (eds.). The Mollusca, volume 10, Evolution. Academic Press, Inc., Orlando, pp. 59–150.

Sankoff, D., R. Cedergren and Y. Abel. 1990. Genomic divergence through gene rearrangement. Methods in Enzymology 183:428–438.

Sankoff, D., G. Leduc, N. Antoine, B. Paquin, B. F. Land and R. Cedergen. 1992. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. Proceedings of the National Academy of Sciences, USA 89:6575–6579.

Sasaki, T., T. Okazaki, M. Muramatsu and R. Kominami. 1987. Variation among mouse ribosomal RNA genes within and between chromosomes. Molecular Biology and Evolution 4(6):594–601.

Scheltema, A. H. 1988. Ancestors and descendents: relationships of the Aplacophora and Polyplacophora. American Malacological Bulletin. 6(1):57–68.

Simpson, L., N. Neckelmann, V. de la Cruz, A. Simpson, J. Feagin, D. Jasmer and K. Stuart. 1987. Comparison of the maxicircle (mitochondrial) genomes of *Leishmania tarentolae* and *Trypanosoma brucei* at the level of nucleotide sequence. Journal of Biological Chemistry. 262(13): 6182–6196.

Smith, A. B. 1989. RNA sequence data in phylogenetic reconstruction: testing the limits of its resolution. Cladistics 5:321–344.

Smith, M. J., D. K. Banfield, K. Doteval, S. Forski and D. J. Kowbel. 1989. Gene arrangement in sea star mitochondrial DNA demonstrates a major inversion event during echinoderm evolution. Gene 76:181–185.

Smith, M. J., D. K. Banfield, K. Doteval, S. Forski and D. J. Kowbel. 1990. Nucleotide sequence of nine protein-coding genes and 22 tRNAs in the mitochondrial DNA of the sea star *Pisaster ochraceus*. Journal of Molecular Evolution 31:195–204.

Smith, M. J., A. Arndt, S. Gorski and E. Fajber. 1993. The phylogeny of echinoderm classes based on mitochondrial gene arrangements. Journal of Molecular Evolution 36: 545–554.

Snyder, M., A. R. Fraser, J. Laroche, K. E. Gartner-Kepkay and E. Zouros. 1987. Atypical mitochondrial DNA from the deep-sea scallop *Placopecten magellanicus*. Proceedings of the National Academy of Sciences, USA 84:7595–7599.

Sogin, M. L. 1991. Early evolution and the origin of eukaryotes. Current Opinion in Genetics and Development 1:457–463.

Solignac, M., M. Monnerot and J. -C. Mounolou. 1986. Concerted evolution of sequence repeats in *Drosophila* mitochondrial DNA. Journal of Molecular Evolution 24:53–60.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98:503–517.

Stuart, K. and J. Feagin. 1992. Mitochondrial DNA of Kinetoplastids. *In*: K. W. Jeon and D. R. Wolstenholme (eds.). Mitochondrial Genomes. International Review of Cytology 141:65–88.

Thomas, W. K., S. Pääbo, F. Villablanca and A. Wilson. 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. Journal of Molecular Evolution 31(2): 101–112.

Tzeng, C. -S., C. -F. Hui, S. -C. Shen and P. C. Huang. 1992. The complete nucleotide sequence of the *Crossostome lacustre* mitochondrial genome: conservation and variations among vertebrates. Nucleic Acids Research 20(18): 4853–4858.

Uhlenbusch, I., A. Mccracken and G. Gellissen. 1987. The gene for the large (16S) ribosomal RNA from the *Locusta migratoria* mitochondrial genome. Current Genetics 11: 621–638.

Vagvolgyi, J. 1967. On the origin of molluscs, the coelom, and coelomic segmentation. Systematic Zoology 16:153–168.

Vahrenholz, C., E. Pratje, G. Michaelis and B. Dujon. 1985. Mitochondrial DNA of *Chlamydomonas reinhardtii*: sequence and arrangement of URF5 and the gene for cytochrome oxidase subunit I. Molecular Genetics 201:213–224.

Vawter, L. and W. M. Brown. 1993. Rates and patterns of base change in the small subunit ribosomal RNA gene. Genetics 134:597–608.

Wainright, P. O. 1993. Monophyletic origins of the metazoa: an evolutionary link with fungi. Science 260(5106):340–342.

Wallace, D. C. 1982. Structure and evolution of organelle genomes. Microbiological Reviews 46(2):208–240.

Walsh, J. 1985. Interaction of selection and biased gene conversion in a multigene family. Proceedings of the National Academy of Sciences, USA 82:153–157.

Warrior, R., and J. Gall. 1985. The mitochondrial DNA of *Hydra attenuata* and *Hydra littoralis* consists of two linear molecules. Archives Science Genéve 38:439–445.

Wheeler, W. C. and R. L. Honeycutt. 1988. Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. Molecular Biology and Evolution 5:90–96.

Wilkinson, G. and A. Chapman. 1991. Length and sequence variation in evening bat D-loop mtDNA. Genetics 128: 607–617.

Wingstrand, K. G. 1985. On the anatomy and relationships of recent Monoplacophora. Galathea Report 16:7–94.

Woese, C. R. and G. E. Fox. 1977. Phylogenetic structure of the prokarotic domain: the primary kingdoms. Proceedings of the National Academy of Sciences, USA 74(11): 5088–5090.

Woese, C. R. 1987. Bacterial evolution. Microbiology Reviews 51(2):221–271.

Wolf, K. and L. Del Guidice. 1988. The variable mitochondrial genome of ascomycetes: organization, mutational alterations, and expression. Advances in Genetics 25:185–308.

Wolstenholme, D. R., J. L. Macfarlane, R. Okimoto, D. O. Clary and J. A. Wahleithner. 1987. Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proceedings of the National Academy of Sciences, USA 84:1324–1328.

Wolstenholme, D. R. 1992a. Genetic novelties in mitochondrial genomes of multicellular animals. Current Opinion in Genetics and Development 2:918–925.

Wolstenholme, D. R. 1992b. Animal mitochondrial DNA: structure and evolution. *In*: K. W. Jeon and D. R. Wolstenholme (eds.). Mitochondrial Genomes. International Review of Cytology 141:173–216.

Zevering, C. E., C. Moritz, A. Heideman and R. A. Sturm. 1991. Parallel origins of duplications and the formation of pseudogenes in mitochondrial DNA from parthenogenetic lizards (*Heteronotia binoei*: Gekkonidae). Journal of Molecular Evolution 33:431–441.

Zwieb, C., C. Glotz and R. Brimacombe. 1981. Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. Nucleic Acids Research 9(15):3621–3640.

The Mitochondrial Genome of *Cepaea nemoralis* (Gastropoda: Stylommatophora): Gene Order, Base Composition, and Heteroplasmy

Jonathan Terrett
Sue Miles
Department of Genetics
Queens Medical Centre
Clifton Boulevard
Nottingham, NG7 2UH, U.K.

Richard H. Thomas
Department of Zoology
The Natural History Museum
Cromwell Road
London, SW7 5BD, U.K.

ABSTRACT

The 14.1 kb mitochondrial genome of the terrestrial gastropod *Cepaea nemoralis* has been cloned and completely sequenced. Sequences coding for 13 proteins and the large and small subunits of ribosomal RNA have been identified. All metazoan mtDNAs examined to date, except for the nematodes *Caenorhabditis elegans* and *Ascaris suum*, and the bivalve mollusc *Mytilus edulis*, contain an ATPase subunit 8 gene. The presence of ATPase 8 in *Cepaea* mitochondrial DNA suggests that this gene has been lost independently from the nematode and bivalve mitochondrial genomes. Commonly genes are encoded on both strands of mtDNA molecules, and this is true of the *Cepaea* mitochondrial genome, but not of the nematodes and *Mytilus*. Base composition is the least biased of any reported metazoan mitochondrial genome. Comparisons are made with other metazoan mitochondrial genomes. The cloned genome has been used to infer the presence of polymorphisms in the size of mitochondrial genomes in *Cepaea* by analysis of the lengths of restriction fragments. These polymorphisms have aided the identification of heteroplasmic snails. A protocol is described that can be used to extract intact gastropod mitochondrial DNA. This DNA is sufficiently free from inhibitors for use in restriction enzyme digestions and for amplification using the polymerase chain reaction.

Key words: Mitochondrial DNA, Mollusca, molecular evolution, heteroplasmy.

INTRODUCTION

The number of completely sequenced mitochondrial genomes is increasing rapidly and the evenness of taxonomic sampling is improving. Sequences of seventeen metazoan mitochondrial genomes are now available from ten vertebrates [six placental mammals (Anderson *et al.*, 1981; Bibb *et al.*, 1981; Anderson *et al.*, 1982; Gadaleta *et al.*, 1989; Arnason *et al* , 1991; Arnason *et al.*, 1992), two fish (Chang & Huang, 1991; Tzeng *et al.*, 1992), a bird (Desjardins & Morais, 1990), and a frog (Roe *et al.*, 1985)], two insects (Clary & Wolstenholme, 1985; Crozier & Crozier, 1993), two sea urchins (Cantatore *et al.*, 1989; Jacobs *et al.*, 1989), and three nematodes (Okimoto *et al.*, 1992; Okimoto *et al* ., 1991). Many other taxa have been sequenced in part and of these the complete gene order is known for three, the mussel, *Mytilus edulis* (Hoffmann *et al.*, 1992), a sea star (Smith *et al.*, 1989) and the codfish, *Gadus morhua* (Johansen *et al.*, 1990). Analyses of these genomes are revealing a host of interesting phenomena (reviewed in Wolstenholme, 1992), as well as providing tools for population and phylogenetic studies (Avise, 1989).

Mitochondrial sequences from molluscs promise to greatly enhance our understanding both of the molecular evolution of metazoan mitochondria and the evolutionary relationships of the molluscs themselves (Boore & Brown, 1994). In this paper we present the order of protein coding genes and ribosomal RNA (rRNA) genes in the mitochondrial genome of *Cepaea nemoralis* and compare it with those of other metazoans. Comparisons of DNA sequence across the phyla can be used to infer mechanisms of mtDNA replication and the mode of base substitution. We make preliminary observations on these points. We also present here a reliable protocol for the extraction of mtDNA from gastropods which is suitable for restriction fragment length polymorphism (RFLP) studies and for polymerase chain reaction (PCR) amplification. Extraction of mitochondrial DNA from molluscs, and in particular terrestrial gastropods, has been problematic (Stine, 1989; J.S. Jones, personal communication); the availability of a simple and reliable method opens up many possibilities for further work on these animals.

MATERIALS AND METHODS

mtDNA extractions: Several protocols for the extraction of mollusc mtDNA have already been published (Skibinski & Edwards, 1987, Stine, 1989). The following

method combines procedures from both of these and can be performed in less than two hours. It has been used successfully and repeatedly on *Cepaea* and *Helix* (Terrett, 1992) and *Littorina* (E. Rumbak, personal communication). Snails are killed by placing at −80°C for 1 hour. They are then defrosted at 50°C for 5 minutes. After removing the snail from its shell, the hepatopancreas is minced thoroughly with a scalpel blade and then homogenised in 3 ml of 0.25M sucrose in TEK (50mM Tris/HCl pH 7.5, 10mM EDTA, 1.5% KCl (w/v)) in a Dounce homogenizer (7ml) with five to ten strokes of the 'A' pestle. The homogenate is transferred to two 1.5ml microfuge tubes. Nuclei and other cell debris are removed by centrifuging at 6,000 rpm (low speed) in a microfuge for two minutes. The supernatants are transferred to two fresh microfuge tubes, and mitochondria collected by centrifugation at 13,000 rpm (high speed) in a microfuge for five minutes. The soft mitochondrial pellets are resuspended in 600 μl of 0.25M sucrose in TEK, and then 600 μl of 1.1M sucrose in TEK is layered beneath. Mitochondria are collected by centrifuging at high speed in a microfuge for 10 minutes. The mitochondrial pellet is then resuspended in 600 μl of STE100 (100mM NaCl, 10mM Tris/HCl pH 8.0, 100mM EDTA) and intact mitochondria are lysed by the addition of 80 μl of 10% Nonidet P–40 followed by gentle shaking. Mitochondrial membranes are removed by centrifugation at high speed in a microfuge for one minute. Mucopolysaccharides are removed by the addition of 100 μl of 5M NaCl to the supernatant followed by 80 μl of CTAB/NaCl (0.7M NaCl / 10% CTAB). This mixture is shaken vigorously before incubating at 65°C for fifteen to thirty minutes. The precipitate that forms contains the mucopolysaccharides and is removed by filling the tube with chloroform, shaking vigorously to form an emulsion, and centrifuging at high speed for three minutes in a microfuge. The mtDNA containing supernatant is removed and subjected to two phenol/chloroform extractions and one chloroform extraction before the nucleic acids are precipitated. This is generally achieved by adding an equal volume of propan–2-ol (20°C) to the aqueous phase, shaking vigorously, and centrifuging immediately at high speed in a microfuge for ten minutes. The yield from the hepatopancreas of a single *Cepaea* is not sufficient for restriction digest fragments to reliably be seen on an agarose gel stained with ethidium bromide (EtBr). However, up to fifteen restriction digests can be visualised after Southern (1975) blotting and detection using digoxigenin (Boehringer Mannheim).

Restriction enzyme digestions and cloning: *C. nemoralis* mtDNA was extracted as above, electrophoresed on a 0.8% EtBr/agarose gel, and open circles were extracted using GenecleanII (Bio 101). This mtDNA was then digested with *Bam*HI and cloned into λGem11 (Promega) using LE392 host. The entire genome was then subcloned into the *Sst*I and *Hind*III sites of pGem7zf(+) (Promega) maintained in *E. coli* strain JM101. All clones were verified as containing *Cepaea* mtDNA by probing against mtDNA extractions. Preparation of plasmid, and phagemid DNAs were as in Sambrook *et al.* (1989).

Sequencing and sequence analysis: DNA sequences were obtained using the dideoxy chain termination method (Sanger *et al.*, 1977) from sets of deletion subclones (Henikoff, 1984) and subsequent extraction of single stranded phagemid DNA (Vieira & Messing, 1987).

Cepaea mitochondrial protein coding genes were identified by the similarity of their inferred amino acid sequences with those of *Drosophila yakuba* (Clary & Wolstenholme, 1985) and humans (Anderson *et al.*, 1981), and in some cases verified by similarities in hydropathy profiles (Kyte & Doolittle, 1982). rRNA genes were identified by the similarities of the DNA sequences of *D. yakuba* and *Cepaea*.

Southern blotting and DNA hybrid detection: Standard Southern (1975) blotting techniques were used to transfer DNA onto nitrocellulose. The entire mitochondrial genome of *C. nemoralis* within its λ vector was used as a probe after labelling with digoxigenin. Labelling and detection procedures were done according to the manufacturer's instructions (Boehringer Mannheim). Blots were washed in 0.1×SSC at 65°C (2 × 15 minutes).

Analysis: Inferred amino acid sequences were aligned with the program CLUSTAL V (Higgins *et al.*, 1992) using the default parameters. Hydrophilicity plots were produced using MacVector (IBI) and additional alignments and analyses were done with DNASTAR.

RESULTS

The entire mitochondrial genome of *Cepaea nemoralis* has been sequenced. At this writing all of the genome has been sequenced on one strand and approximately 4,000 bases of the 14,099 bp genome have been confirmed by sequencing both strands (Terrett, 1992). The sequence has enabled the arrangment of the protein and rRNA genes to be established and the base composition to be analysed.

Southern blots of *Cepaea* mtDNA that had been digested with *Kpn*1 show polymorphism in the size of the mitochondrial genome and heteroplasmy.

DISCUSSION

Gene Content

The 'full' complement of thirteen protein coding genes has been identified in the *Cepaea* sequence, i.e., cytochrome b (cytb), cytochrome oxidase subunits I to III (COI, COII, and COIII), NADH dehydrogenase subunits 1–6 and 4L (ND 1–6, 4L), ATP synthase subunits 6 and 8 (ATPase 6 and 8) and two rRNA genes (12S and 16S). Many of the genes were easily identified by comparing the *Cepaea* sequences with the *Drosophila* sequences (Clary & Wolstenholme, 1985). However, although conserved amino acid domains were identified in ND6, ND4L, and ATPase 8 (Figures 1–3), these were not long

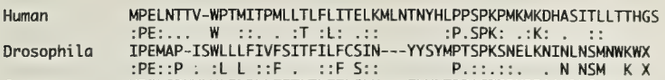


Figure 1. Amino acid alignment of the ATPase 8 protein sequences from *Homo sapiens*, *Drosophila yakuba* and *Cepaea nemoralis*. Identical amino acids are indicated by single letter codes between the sequences and conservative substitutions (French & Robson, 1983) by colons.

enough to be confident of their identification. The similarity of the hydrophilicity profiles of the proposed *Cepaea* proteins to those of other organisms (Figure 4) helped to verify the identity of these genes.

Most metazoan mtDNAs encode an ATPase subunit 8. The exceptions to date are the nematodes *Caenorhabditis elegans* and *Ascaris suum* (Okimoto *et al.*, 1992), and the bivalve mollusc *Mytilus edulis* (Hoffmann *et al.*, 1992). The presence of ATPase 8 in *Cepaea* mitochondrial DNA suggests that this gene has been lost independently from the bivalve and nematode mitochondrial genomes.

Has the function of ATPase 8 been taken over by some other protein in mussels and nematodes? Have there been two independent losses of the ATPase 8 gene from mitochondrial lineages leading to mussels and nematodes? Did a common ancestor of molluscs and nematodes relocate the mitochondrial ATPase 8 gene to the nucleus, only for the *Cepaea* lineage to move it back again? More sequences from molluscan taxa are required to answer these questions.

The *Cepaea* ATPase 8 gene is adjacent to, but not overlapping with, ATPase 6. However, in *Cepaea* ATPase 8 and ATPase 6 are in the opposite orientation from *Drosophila*, vertebrates, and echinoderms, where the two genes overlap in the order ATPase 6–ATPase 8. In an attempt to explain the continued existence of mitochondrial genomes, von Heijne (1986) suggested that ATPase 8 has not moved to the nuclear genome because the overlap with ATPase 6 cannot be overcome. ATPase 6 remains encoded by mtDNA because hydrophobic domains within the protein would act as signal sequences causing it to be incorrectly transported (i.e. not to the mitochondrial membrane). The absence of ATPase 8

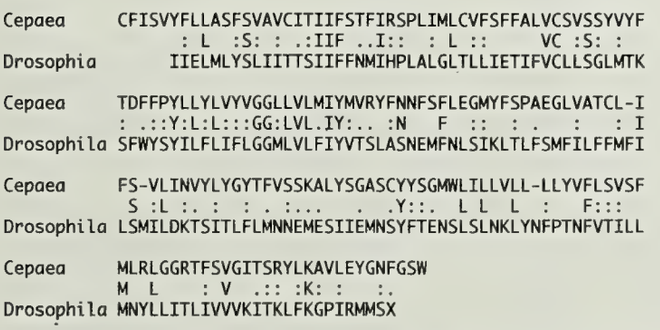


Figure 2. Amino acid alignment of the ND6 proteins of *Cepaea nemoralis* and *Drosophila yakuba*. Similarities are as in Figure 1.

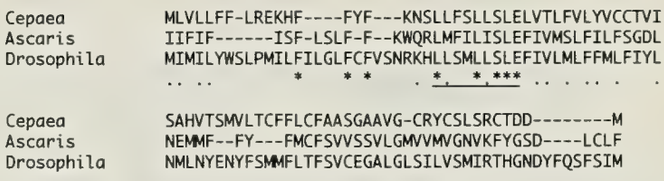


Figure 3. Amino acid alignment of the ND4L proteins of *Cepaea nemoralis*, *Drosophila yakuba*, and *Ascaris suum*. Identical amino acids are indicated by asterisks, conservative substitutions by dots, and a conserved region is underlined.

from *Mytilus* and nematodes, and its relocation in *Cepaea* cast doubts upon the necessity of the overlap.

Gene order

When the first few mitochondrial genomes were sequenced the identical gene order seen in placental mammals (e.g. Anderson *et al.*, 1981) and a frog (Roe *et al.*, 1985), plus the possibility of deriving the gene order of the protein and rRNA genes of a sea urchin (Cantatore *et al.*, 1989) from that of a vertebrate with only two rearrangements (Wolstenholme, 1992), gave the impression that gene order was likely to be very conservative within phyla. The strikingly different gene orders of the bivalve *Mytilus* and the pulmonate gastropod *Cepaea* show that mitochondrial gene order is not necessarily at all conservative.

Figure 6 shows a linearised comparison of the genomes of *Mytilus* and *Cepaea*. The lack of similarity in the gene orders is striking. The *Cepaea* genome, unlike *Mytilus*, contains genes encoded on both DNA strands. Inferring simple inversion events in either organism does not increase the similarity. *Cepaea* and vertebrates share some gene boundaries but the transcriptional orders are all different, suggesting the similarity does not result from shared ancestry. The difference in gene order between these two molluscs is by far the greatest reported within any phylum.

The location of genes on both strands of the *Cepaea* genome is particularly interesting because each set of protein genes on each strand has a rRNA gene between them (Figure 5). rRNA sequences have been inferred to function during initiation of transcription of mtDNA (Montoya *et al.*, 1982). There are very few nucleotides in the *Cepaea* mitochondrial genome that are not assigned to genes (Terrett, 1992). Previously sequenced metazoan mtDNAs contain sequences involved in replication and transcription, e.g. the D-loop in vertebrates (Clayton, 1992), and A+T rich regions in insects (Clary & Wolstenholme, 1985). It will be interesting to see where transcription and replication are initiated in the tightly packed genome of *Cepaea*.

Base composition and codon usage

The base composition of mtDNAs was first shown to be biased when Brown (1981) separated the two strands of human mtDNA on a cesium chloride gradient by virtue

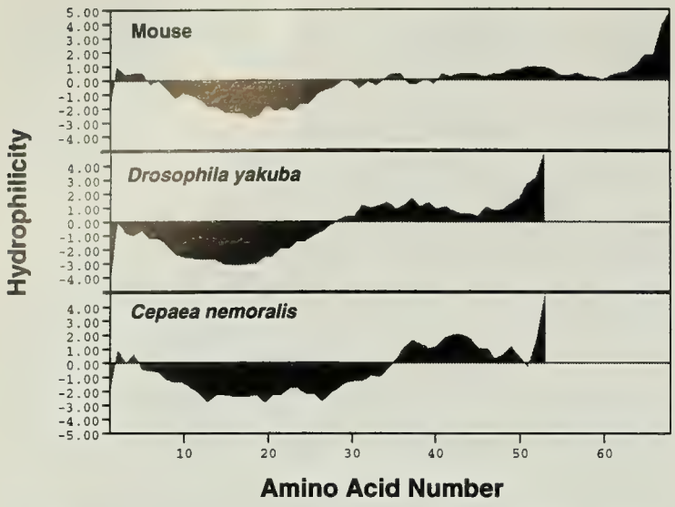


Figure 4. Hydrophilicity profiles of ATPase 8 proteins showing the characteristic hydrophobic amino end and hydrophilic carboxy end.

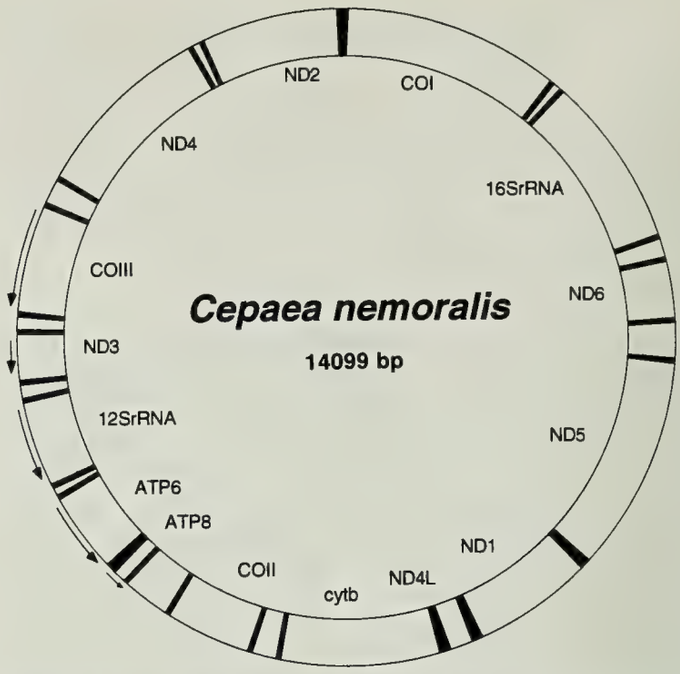


Figure 5. The mitochondrial genes of *Cepaea nemoralis* in their native circular arrangement. Most genes are transcribed in the clockwise direction. The minor strand encodes five genes transcribed in the other direction (small arrows).

of their different base compositions and thus their different molecular weights. This bias has now been characterised in vertebrates as being caused by a bias towards G and T on one strand, and towards C and A on the other. The mtDNA's of nematodes and *Drosophila* show such a large bias towards A and T that the strands do not differ significantly in density. Table 1 shows the base compositions at the third positions of codons for the *Cepaea* protein genes, separated by coding strand. A and T contents are similarly high on both strands and there appears to be a slight bias against G on the minor (fewer genes) strand.

A further indication of a small bias in the mtDNA of *Cepaea* can be seen by examining the frequencies of the two families of leucine codons. They are grouped as TTR (R represents G or A) and CTN (N is any base). Thus there are two TTR codons and four CTN codons. If codon usage is independent of base composition then we would expect roughly twice as many CTN codons than TTR codons. However, in the A+T rich genome of *A. suum*, the ratio of TTR:CTN codons is 6.7:1 (Okimoto *et al.*, 1992). Within the CTN group in *A. suum* the T bias is obvious as 57 of 67 CTN codons are CTT. In *Cepaea* fairly even use of these codons can be seen, the TTR:CTN ratio being 0.81:1 for all proteins (the ratio for only the minor strand is remarkably similar; 0.75:1) so that the slight bias towards A and T is all that is evident. This is reinforced by the CTN group as 206 codons are CTA/T and 120 codons are CTG/C. The mtDNA of *Mytilus* shows a greater bias (especially against C at the third position of codons on the sense strand) but this bias is very small compared with those of the vertebrates.

Polymorphism in genome size

As mtDNA's are studied more at the population level, and thus more individuals are screened, we are becoming increasingly aware of polymorphisms in the sizes of mitochondrial genomes within species. At the extreme, variation can occur within one generation (Cook & Zouros, 1994). Figure 7 shows the existence of a size polymorphism in a *Cepaea* population. If this polymorphism is as rapidly fluctuating as that in scallops (Gjetvaj *et al.*, 1992), our ability to use mtDNA restriction fragment length polymorphisms to infer the population structure of *Cepaea* will be limited. However, the sequence can be used to design primers to amplify and sequence regions of the mitochondrial genome not subject to the size variation.

That the mitochondrial genomes of *Cepaea* and *Mytilus* cannot be aligned with respect to gene order shows the potential within the molluscs of inferring relatedness from gene orders (Boore & Brown, 1994). As with all comparisons at the molecular level, plateaus will be reached above which comparisons become meaningless. This point is demonstrated with the *Mytilus* and *Cepaea* gene orders. There is also great opportunity to study the

Table 1. Percent base composition at codon third positions in the mitochondrial protein genes of *Cepaea nemoralis*. The major strand encodes nine proteins and the minor strand encodes four proteins.

	Percent base composition	
Base	Major strand	Minor strand
G	19	13
A	27	23
T	36	38
C	18	26

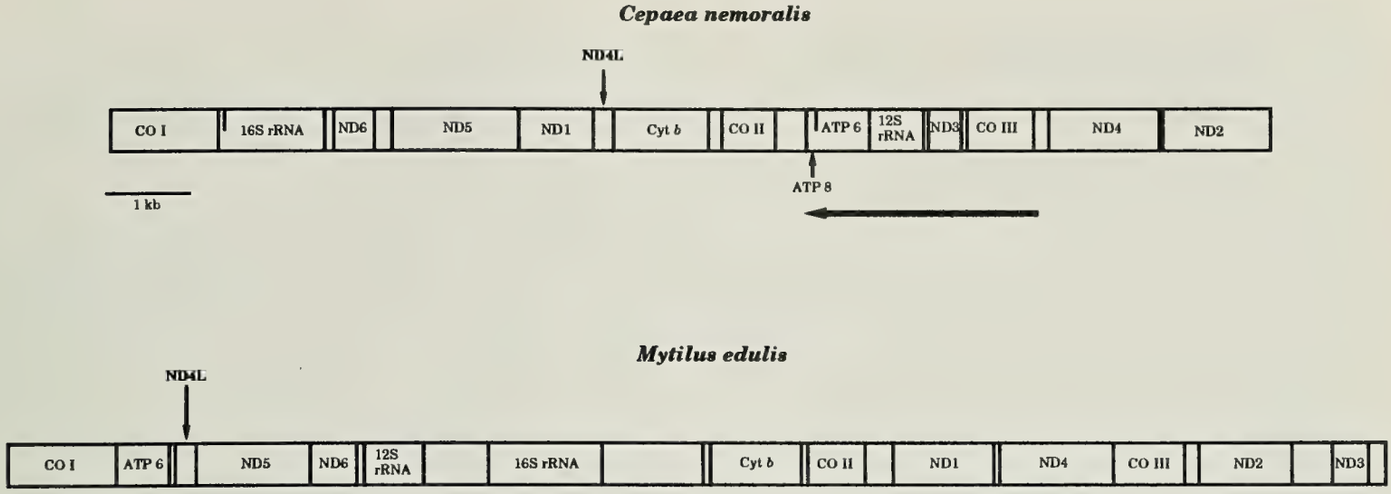


Figure 6. Linear representation of the mitochondrial genomes of *Cepaea nemoralis* and *Mytilus edulis* (Hoffmann *et al.*, 1992). Arrows indicate the DNA strand on which genes are encoded. Unlabelled sections of the genomes contain tRNAs and noncoding sequences. There is no obvious way to derive one gene order from the other.

evolution of mtDNA at a fine level within the molluscs. The polymerase chain reaction (PCR) could be employed to speed up the production of these data.

ACKNOWLEDGEMENTS

We thank the SERC and the NERC for financial support, and David Skibinski for useful discussions on extraction procedures.

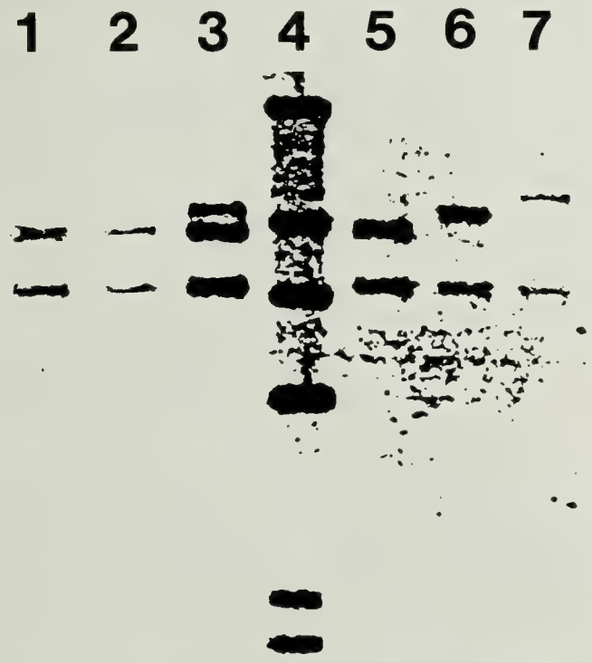


Figure 7. *Kpn*I restriction digests of mtDNAs isolated from individual snails. Lanes 1–3 were collected from Derbyshire and 5–7 from Lincolnshire. Lane 4 is λ DNA cut with *Hind* III. Lanes 1, 2, and 5 show fragments expected from the cloned sequences. Thus lanes 6 and 7 contain a larger than expected fragment. Lane 3 shows the expected fragments plus the larger one present in lane 6 and this snail was probably heteroplasmic.

LITERATURE CITED

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. de Bruijn., A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J .H. Smith, R. Staden, and I. G. Young. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457–465.

Anderson, S., M. H. L. de Bruijn, A. R. Coulson, I. C. Eperon, F. Sanger, and I. G. Young. 1982. Complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. Journal of Molecular Biology 156:683–717.

Arnason, U., A. Gullberg, and B. Widegren. 1991. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. Journal of Molecular Evolution 33:556–568.

Arnason, U. and E. Johnsson. 1992. The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina*. Journal of Molecular Evolution 34:493–505.

Avise, J. C. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution 43: 1192–1208.

Avise, J. C., B. W. Bowen, T. Lamb, A. B. Meylan, and E. Bermingham. 1992. Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the Testudines. Molecular Biology and Evolution 9:457–473.

Bibb, M. J., R. A. van Etten, C. T. Wright, M. W. Walberg, and D. A. Clayton. 1981. Sequence and gene organization of mouse mitochondrial DNA. Cell 26:167–180.

Boore, J. L. and W. M. Brown. 1994. Mitochondrial Genomes and the Phylogeny of Mollusks. *In:* Harasewych, M. G. and S. Tillier (eds.) Molecular Techniques and Molluscan Phylogeny: Proceedings of a Symposium held at the 11th International Malacological Congress, Siena. The Nautilus, Supplement 2:61–78 (This volume).

Brown, W. M. 1981. Mechanisms of evolution in animal mitochondrial DNA. Annals of the New York Academy of Sciences 361:119–134.

Cantatore, P., M. Roberti, G. Rainaldi, M. N. Gadaleta, and C. Saccone. 1989. The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial ge-

nome of *Paracentrotus lividus*. Journal of Biological Chemistry 264:10965–10975.

Chang, Y. S., and F. L. Haung. 1991. *Cyprinus carpio* complete mitochondrial genome DNA. Unpublished. GenBank Accession X61010.

Clary, D. O., and D. R. Wolstenholme. 1985. The mitochondrial DNA molecule of *Drosophila yakuba* nucleotide sequence, gene organization, and genetic code. Journal of Molecular Evolution 22:252–271.

Clayton, D. A. 1992. Transcription and replication of animal mitochondrial DNA. International Review of Cytology 141: 217–232.

Cook, D. I. and E. Zouros. 1994. The highly variable and highly mutable mitochondrial DNA molecule of the Deep Sea Scallop *Placopecten magellanicus*. *In:* Harasewych, M. G. and S. Tillier (eds.) Molecular Techniques and Molluscan Phylogeny: Proceedings of a Symposium held at the 11th International Malacological Congress, Siena. The Nautilus, Supplement 2:85–90 (This volume).

Crozier, R. H., and Y. C. Crozier. 1993. The mitochondrial genome of the honeybee, *Apis mellifera*: complete sequence and genome organisation. Genetics 133:97–117.

Desjardins, P. and R. Morais. 1990. Sequence and gene organisation of the chicken mitochondrial genome: a novel gene order in higher vertebrates. Journal of Molecular Biology 212:599–634

French, S., and B. Robson. 1983. What is a conservative substitution? Journal of Molecular Evolution 19:171–175.

Gadaleta, G., G. Pepe, G. De Candia, C. Quagliariello, E. Sibisa, and C. Saccone. 1989. The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparitive analysis between vertebrates. Journal of Molecular Evolution 28:497–516.

Gjetvaj, B., D. I. Cook, and E. Zouros. 1992. Repeated sequences and large-scale size variation of mitochondrial DNA: a common feature among scallops (Bivalvia: Pectinidae). Molecular Biology and Evolution 9:106–124.

von Heijne, G. 1986. Why mitochondria need a genome. FEBS Letters 198:1D4.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351–359.

Higgins, D. G., A. J. Bleasby, and R. Fuchs. 1992. CLUSTAL - V - Improved software for multiple sequence alignment. CABIOS 8:189–191.

Hoffmann, R. J., J. L. Boore, and W. M. Brown. 1992. A novel mitochondrial genome organisation for the blue mussel, *Mytilus edulis*. Genetics 131:397–412.

Jacobs, H. T., D. J. Elliott, V. B. Math, and S. Farquharson. 1988. Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. Journal of Molecular Biology 202:185–217.

Johansen, S., P. H. Guddal, and T. Johansen. 1990. Organisation of the mitochondrial genome of Atlantic Cod. Nucleic Acids Research 18:411–419.

Kyte, J. and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 157:105–132.

Montoya, J., T. Christianson, D. Levens, M. Rabinowitz, and G. Attardi. 1982. Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. Proceedings of the National Academy of Sciences, U.S.A. 79:7195–7199.

Okimoto, R., H. Chamberlin, J. L. Macfarlane, and D. R. Wolstenholme. 1991. Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (*Meloidogyne*): nucleotide sequences, genome location, and potential for host-race identification. Nucleic Acids Research 19: 1619–1626.

Okimoto, R., J. L. Macfarlane, D. O. Clary, and D. R. Wolstenholme. 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics 130:471–498.

Roe, B.A., D.-P. Ma, R. K. Wilson, and J F.-H. Wong. 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. Journal of Biological Chemistry 260:9759–9774.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: a laboratory manual. Second edition. Cold Spring Harbor Laboratory Press.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Sciences, U.S.A. 74:5463–5467.

Skibinski, D. O. F. and C. A. Edwards. 1987. Mitochondrial DNA variation in marine mussels (*Mytilus*). Proceedings of the World Symposium on selection, hybridisation and genetic engineering in aquaculture, Bordeaux 27 –30 May 1986. Vol. 1, Berlin.

Smith, M.J., D.K. Banfield, K. Doteval, S. Gorski, and D.J. Kowbel. 1989. Gene arrangement in sea star mitochondrial DNA demonstrates a major inversion event during echinoderm evolution. Gene 76:181–185.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98:503–517.

Stine, O. C. 1989. *Cepaea nemoralis* from Lexington, Virginia: The isolation and characterization of their mitochondrial DNA, the implications for their origin and climatic selection. Malacologia 30:305–315.

Terrett, J. A. 1992. The mitochondrial genome of *Cepaea nemoralis*. Ph.D. Thesis. University of Nottingham.

Tzeng, C.-S., C.-F. Hui, S.-C. Shen and P. C. Huang. 1992. The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome - conservation and variations among vertebrates. Nucleic Acids Research 20:4853–4858.

Vieira, J. and J. Messing. 1987. Production of single stranded plasmid DNA. Methods in Enzymology 153:3.

Wolstenholme, D. R. 1992. Genetic novelties in mitochondrial genomes of multicellular animals. Current Opinions in Genetics and Development 2:918–925.

The Highly Variable and Highly Mutable Mitochondrial DNA Molecule of the Deep Sea Scallop *Placopecten magellanicus*

Douglas I. Cook[1]
[1] Marine Gene Probe Laboratory
Department of Biology
Dalhousie University
Halifax, N.S., Canada, B3H 4J1

Eleftherios Zouros [1,2]
[2] Department of Biology
University of Crete and
Institute of Marine Biology of Crete
711 10 Iraklion, Crete, Greece

ABSTRACT

Whereas in the vast majority of animals the mitochondrial DNA (mtDNA) molecule rarely exceeds 20 kb (kilobases) in length and is size-invariable among conspecific individuals, in several invertebrates, fish, amphibians and reptiles the length of this molecule can be much larger and may vary within and among individuals. An extreme case is presented by the deep-sea scallop *Placopecten magellanicus*, where the molecule is, on average, 36 kb and may vary from 31 to 42 kb. Most of this variation is due to arrays of tandemly repeated sequences whose copy number varies among molecules. Such arrays (each consisting of a different repeated sequence) occur in more than one region of the mtDNA molecule. Once present, these arrays allow for frequent errors to occur during DNA replication, with the result that new size variants appear at a very high rate. The resulting mtDNA size polymorphism may be instrumental for long-term evolutionary phenomena (such as the rearrangement of genes around the molecule), but, because of its rapid turn-over, does not provide useful information for taxonomic studies.

Key Words: mtDNA size variation, scallops, mtDNA evolution.

INTRODUCTION

Several recent reviews (e.g. Harrison, 1989; Avise, 1991; Meyer, 1994) have dealt with the advantages of animal mitochondrial DNA (mtDNA) as a tool for taxonomic and population genetics studies. Briefly, these include the ease with which the mtDNA can be separated from nuclear DNA, the property of homoplasmy (i.e. that all mtDNA molecules of an individual organism are identical), the accumulation of base substitutions at a faster rate than in coding single copy nuclear DNA, and the maternal transmission which is particularly useful for studies of population structure in species whose distribution is mainly affected by female dispersion.

With the exception of maternal inheritance (which applies to the mtDNA molecule as a whole), these properties of the animal mtDNA appear to be true for nucleotide variation, but not for variation caused by changes in the length of the molecule. Length variations in animal mtDNA were originally considered to be either very rare or of negligible size (Attardi, 1985). This still appears to be the case for homeotherms. In many species of invertebrates, fish, amphibians and reptiles, however, the length of the mtDNA molecule can vary substantially from individual to individual, and the same individual may carry mtDNA molecules of different lengths (for review see Moritz *et al.*, 1987). It became clear from the first studies of length polymorphism that this type of variation is different from single nucleotide substitution in several ways. Whereas nucleotide substitutions may occur more or less randomly around the molecule, length polymorphism is localized in (or near) a specific region of the genome, the region that controls the initiation of transcription. Also, whereas nucleotide heteroplasmy (the presence of two or more types of mtDNA molecules in the same organism) is rare, length heteroplasmy is common. In fact, many length variants in several species have been observed only in the state of heteroplasmy (Willis, 1987; Bentzen *et al.*, 1988).

There are two main types of mtDNA length variation in animals. One is due to single duplications of part of the molecule. The other is due to repeated sequences, whose copy number vary among molecules. The level of variation of the latter type [which is akin to the VNTR (variable number of tandem repeats) polymorphism of nuclear DNA] is generally much higher than that of the first.

These differences between base substitution variation and length variation suggest that the two types of mtDNA polymorphism have quite different evolutionary dynamics and, thus, can have different uses in population studies. This communication provides a short description of size polymorphism of an exceptionally large mtDNA molecule, that of the deep-sea scallop, and, also, reports original observations about the rate with which new size variants appear in this molecule.

AN OVERVIEW OF THE mtDNA OF THE DEEP SEA SCALLOP, *Placopecten magellanicus* (Gmelin, 1791)

The mtDNA of the deep-sea scallop remains the largest metazoan mtDNA reported (Snyder *et al.*, 1987), even though exceptionally large mtDNA molecules were reported in the nematode *Romanomermis culicivorax* (Powers *et al.*, 1986) and bark weevils of the genus *Pissodes* (Boyce *et al.*, 1989). Figure 1 is a simplified version of the restriction map showing the regions of the molecule at which size variation has been observed. Most of the variation is contained within a part of the genome defined by the *Eco*RI #1 site (arbitrarily positioned at the 12 o'clock point of the circular molecule) and the first *Sph*I site clockwise from the *Eco*RI #1 site. Originally, it was thought that this part of the molecule contained two adjacent yet independently varying subregions, named "locus I" and "locus II". Locus I consists of tandemly repeated copies of a 1449 bp (base pair) sequence, the number of copies varying from two to eight. The molecule shown in figure 1 contains three such repeats. The molecular basis of the polymorphism at locus I was studied by LaRoche *et al.* (1990), and its distribution in natural populations by Fuller (1991). Locus II is defined by the *Eco*RI #1 and the first *Kpn*I site clockwise from *Eco*RI #1. It is approximately 1000 bp in length and varies in size by deletions or duplications of as much as 450 bp. Subsequent analysis has shown that a large part of locus II bears high sequence similarity with the repeated sequence of locus I, so that the two kinds of size variation cannot be treated independently from each other. Locus III is positioned at the 8 o'clock region of the molecule, separated by more than 10 kb from locus I. A large number of size variants occur at this locus, varying from each other by increments of less than 100 bp.

The occurrence of at least two independent size polymorphisms at two disjoint regions of the mtDNA is a rare phenomenon. In addition to the deep sea scallop (Fuller & Zouros, 1993), it has been observed in the Icelandic scallop, *Chlamys islandica* (Gjetvaj *et al.*, 1992). It generates very complex patterns of heteroplasmy, so that each animal may possess a different mtDNA profile. The high level of heteroplasmy introduces, also, an additional level of mtDNA variation, intra-individual variation, that does not exist for the nuclear DNA. Because the state of variation at any given level is the outcome of the interplay between deterministic (e.g. selection) and stochastic (e.g. random drift) forces at that particular level, the presence of intra-individual variation raises the possibility of molecular (among molecules within the individual) selection for mtDNA.

The apportionment of mtDNA variation at different nested levels (among molecules within the individual, among individuals within the population, among populations within the species) has been the subject of several theoretical and empirical studies (Birky *et al.*, 1983; Birky *et al.*, 1989; Rand & Harrison, 1989; Arnason & Rand, 1992). The latter have suggested that the largest component of mtDNA size variation is the among-individuals (within populations). Assuming neutrality, it is easy to see that the ratio of "within individuals" to "among individuals" components of variation depends on the rate with which new variants arise in the population (the mutation rate for length polymorphism) and on the rate with which the resulting state of heteroplasmy "decays" into homoplasmy through a mechanism of stochastic assortment of the constituent mtDNA molecules (Solignac *et al.*, 1984; Rand & Harrison, 1986). It is obvious that the higher the mutation rate is, the higher will be the "within individuals" component. The high rate of heteroplasmy of scallop mtDNA suggests that the rate of length mutation is very high in this molecule. No estimates of mtDNA length mutation rates exist, however, in the literature. We, therefore, attempted to obtain such an estimate in the scallop.

HIGH MUTATION RATE FOR LENGTH POLYMORPHISM

A direct estimation of the rate with which new length variants appear in the mtDNA molecule of the deep sea scallop was attempted by examining the mtDNA of offspring from pair matings and comparing it to that of the parents. We have confined our study to one family and examined locus I and locus II, but not locus III. Total mtDNA was extracted from one year-old progeny according to the techniques described in Fuller (1991) and Zouros *et al.* (1992a). Ten μg of DNA was digested with *Eco*RI or a combination of *Sal*I/*Stu*I. The resulting fragments were separated on 0.7% agarose, transferred to nylon membrane and hybridized either to a *Pst*I/*Pst*I clone of the repeat element of locus I (LaRoche *et al.*, 1990) or to an *Eco*RI/*Pst*I clone of the fragment contained between the *Eco*RI #1 and the first *Pst*I site clockwise from *Eco*RI #1 (Fig. 1).

Table 1 presents the results in a way that does not require familiarity with the molecular data. The "genotype" for each individual animal refers to the state of locus I and locus II. The characteristic number (varying from 3 to 6) for each genotype indicates the number of copies of the repeat element at locus I. A "minus" superscript indicates a deletion at locus II. Individuals with two numbers are heteroplasmic.

The mother was apparently homoplasmic for the type of molecule that contains four copies of the repeat element at locus I and carries no deletions or insertions at locus II. This is the most common "genotype" in all populations of this species (Fuller, 1991). The father was heteroplasmic, with the 5^- type representing the minority of mtDNA molecules in this individual. In all, 87 offspring were scored. All of them had the mother's mtDNA type, either exclusively or in conjunction with another type. Seventy-four offspring had only the mother's type. Of the remaining thirteen, eleven had an additional type that was completely new (found in neither parent). In two offspring (those of the genotype 4/3) the non-maternal type is similar to one of the two paternal types.

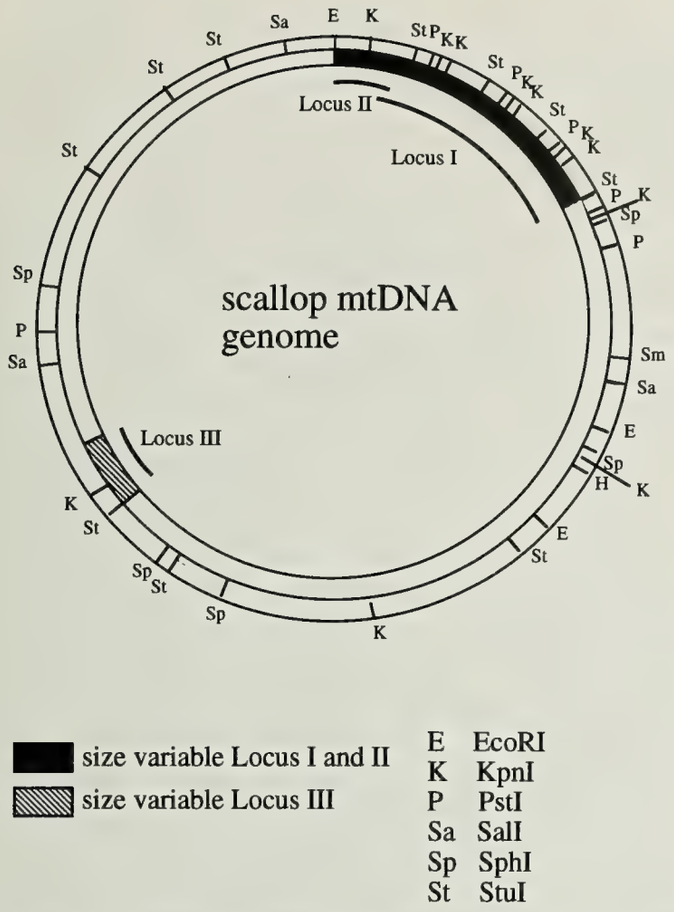


Figure 1. A restriction map of the mitochondrial DNA of the deep sea scallop *Placopecten magellanicus*. The molecule shown has three copies of the repeated sequence at locus I and is of the standard type (i.e. it does not contain deletions or insertions) at locus II. Its size is approximately 36 kb. The *Eco*RI site placed at the 12 o'clock position is referred to in the text as *Eco*RI #1 and is used as a reference point.

Theoretically this could be a case of paternal transmission of mtDNA, a phenomenon known to be common in another bivalve, the blue mussel *Mytilus* (Zouros *et al.*, 1992b). Yet, given that non-parental types (6, 5 and 4$^-$) have appeared among offspring, it is equally or even more probable that the type 3 molecule in the two offspring has resulted from a mutational event in the mother's mtDNA rather than inherited from the father.

Assuming that the non-maternal types in all thirteen progeny represent independent mutational events, the mutation rate is 15%, a very high rate. If the type 3 molecule is attributed to paternal inheritance, the rate would reduce to 12.5%, a trivial change from 15%. In seven of the thirteen progeny the non-maternal type resulted from a mutational event at locus I (types 3, 5 and 6) and in six from a mutational event at locus II (type 4$^-$). Thus the mutation rate is about the same (approximatelly 7.5 %) at each of these loci. It is, however, possible that only one mutational event in the mother's germ line produced the 4$^-$ type, that this type was amplified stochastically in the germ cells and that copies were transmitted to all six progeny. The same argument

Table 1. The genotypes of the parents and of eighty-seven offspring at locus I and locus II of the mtDNA of the deep-sea scallop *Placopecten magellanicus* (for details refer to Figure 1).

Individual	Genotype	Comment
Mother	4	Four copies of the repeat element at locus I; no detectable deletions or insertions at locus II
Father	3/5$^-$	Heteroplasmic. Most common molecule has three copies of the repeat element at locus I; no detectable deletions or insertions at locus II. Less common molecule has five copies of the repeat element at locus I and a 443 bp deletion at locus II.
74 offspring	4	Like mother.
6 offspring	4/4$^-$	Heteroplasmic. One type of molecule has four copies of the repeat element at locus I; no detectable deletions or insertions at locus II. The other type has four copies of the repeat element at locus I and a 443 bp deletion at locus II.
4 offspring	4/5	Heteroplasmic. One type of molecule has four copies of the repeat element at locus I, the other type has five. No detectable deletions or insertions at locus II in either molecule.
2 offspring	4/3	Heteroplasmic. One type of molecule has four copies of the repeat element at locus I, the other three. No detectable deletions or insertions at locus II in either molecule.
1 offspring	4/6	Heteroplasmic. One type of molecule has four copies of the repeat element at locus I, the other six. No detectable deletions or insertions at locus II in either molecule.

can be made for the four offspring with the genotype 4/5, and the two offspring with the genotype 4/3. Under this logic, the mutation rate becomes 4/87 or 4.5%, still a high mutation rate.

MOLECULAR CHARACTERIZATION OF SPONTANEOUS LENGTH MUTATIONS

To further characterize the length variants that appeared in thirteen offspring, the DNA of these offspring and that of the parents was digested with a combination of *Sal*I and *Stu*I enzymes and hybridized to a clone containing the *Eco*RI/*Pst*I part of locus II. From Figure 1 it can be seen that this double digestion will decompose the fragment defined by the first *Sal*I site counter-clockwise

from *Eco*RI #1 and the first *Sal*I site clockwise from *Eco*RI #1 into three size-fragments: the "left flanking region" starting from the first *Sal*I site and ending at the *Stu*I site of the first repeated element of locus I, a number of *Stu*I/*Stu*I fragments of the size of the repeated element, and the "right flanking region" starting from the *Stu*I site of the last repeated element and ending at the second *Sal*I site. All these fragments will react with the *Eco*RI/*Pst*I clone, because this clone contains locus II and a large part of the repeated element of locus I. The seven offspring with the genotypes 4/6, 4/5 and 4/3 produced patterns identical to their mother, implying that the non-maternal type in these offspring resulted from spontaneous insertions of one or two complete elements (for genotypes 4/5 and 4/6, respectively) or from spontaneous excisions of a complete element (for genotypes 4/3). In contrast, the "left flanking region" in all offspring with the phenotype $4/4^-$ was represented by two bands, one identical and one shorter than the corresponding flanking region of the mother, implying that a deletion has occurred in this region. Interestingly, the shorter left flanking region of these offspring was identical with the corresponding region of the minor molecule of the father, implying that the minor component of the father's mtDNA also contains the same sized deletion. In the father this deletion at locus II is followed by five repeats at locus I, whereas in the offspring it is followed by four.

To find out whether the deletion was exactly the same in all six offspring with the phenotype $4/4^-$, the region between *Eco*RI #1 and the first *Kpn*I site was cloned from a molecule that did not carry the locus II deletion and sequenced. Primers were designed from this sequence, and the corresponding region was amplified by PCR (polymerase chain reaction) from several individuals. In all $4/4^-$ offspring and in the father the amplification produced two products, a "long" product resulting from molecules of type 4 (in the offspring) or type 3 (in the father), and a "short" product resulting from molecules of type 4^- (in the offspring) or of type 5^- (in the father). The short product was sequenced and found to be identical in all cases. It differed from the long product by the same deletion of 443 bp. The start and end points of the deletion are marked by a sequence of 86 nucleotides that occurs twice within the *Eco*RI/*Kpn*I fragment.

DISCUSSION

The main results of this study can be summarized as follows:

1. An exceptionally large number (15 %) of offspring from a pair-mating were found to carry in the state of heteroplasmy a new length variant of mtDNA not present in the mother. It can be argued that these new types were in fact present in the mother, but only in minute amounts and that they were disproportionately transmitted or amplified in a few offspring. This would mean that the mother contained at least four minor types of DNA, in addition to the major 4 type. In their surveys of natural populations, Fuller (1991) scored about 350 individuals and Zouros *et al.* (1992a) an additional 250 for heteroplasmy at locus I. The rate of heteroplasmy was found to vary from 10% to 20%, but very rarely (0.4%) was an individual observed with more than two types of mtDNA molecules. On the basis of this observation, the possibility that the mother of the family we examined was pentaplasmic (i.e. it contained five types of molecules) appears unlikely.

2. All non-maternal types seen in the offspring were previously known to exist in natural populations of the species. Thus, length mutations cannot be of any arbitrary size, but rather occur in "quantum" steps whose length corresponds to the length of a repeated sequence or to the length of a unique sequence flanked by repeated sequences.

These results provide direct support for several hypotheses regarding the origin and fate of length variation of mtDNA. Indirect evidence for high mutation rate was deduced from the high frequency of length heteroplasmy in many species (Zouros *et al.*, 1992a, Brown *et al.*, 1992; Arnason & Rand, 1992). More direct evidence for high mutation rate was obtained in the nematode *Romanomermis culicivorax* (Hyman & Slater, 1990), where novel mtDNA forms were observed in lineages separated by less than two hundred generations. In these experiments it was not possible, however, to obtain even an approximate estimate of mutation rate, because it was not possible to monitor the appearance or loss of new mtDNA variants at discrete generation intervals.

Length mutations apparently result from errors during the replication of the molecule, and these errors are, in turn, mediated by the presence of repeated sequences. Possible mechanisms involve replication slippage (Hauswirth *et al.*, 1984), competitive strand displacement (Buroker *et al.*, 1990), or recombination (Rand & Harrison, 1989), even though the latter appears less likely. Among the thirteen mutational events we have observed, six (those leading to the 4^- deletion at locus II) involved the excision of a unique 443 bp sequence flanked at both sides by an identical 86 bp sequence. If these two identical sequences pair with each other at a certain stage of DNA replication, the in-between unique sequence would "loop out" and could be excised, thus leading to a deletion event. Two other mutational events involved the shortening of the array of repeated elements at locus I by one repeat (the type 3 molecule). This can also result from an excision of a copy that looped out as a result of the pairing of the two copies flanking, at both sites, the excised. copy. Finally, one event involved the insertion of one copy of the repeated element (the type 5 molecule) and another event involved the insertion of two copies (the type 6 molecule). These events can be explained by assuming that a segment of DNA looped out (as a result of the pairing of two copies separated by one or two other copies) and was replicated.

Whatever the exact molecular mechanism is for the spontaneous changes of the length of the mtDNA mol-

ecule, it is clear that these changes require the presence of repeated sequences and that, once this requirement is met, they will occur at a very high rate. It is less clear whether mutations of this kind are neutral to the forces of natural selection acting on the individual, even though there appears to be indirect evidence in support of neutrality (Zouros *et al.*, 1992a). Intra-individual molecular selection appears to favor smaller molecules over longer ones (Rand & Harrison, 1986; Brown *et al.*, 1992). Such selection will eliminate molecules of larger size, unless it is countered by a biased mutation rate favoring the generation of longer molecules from shorter ones. In our data no such bias is evident, but the number of mutations we have observed is not large enough for a meaningful test of this hypothesis.

The high mutation rate generates a large amount of intra-individual and inter-individual variation, which may be useful for the study of the evolution of the molecule itself, and may also provide the building blocks for long-term evolutionary changes, such as gene re-arrangements, known to have occurred on multiple occasions (Hoffmann *et al.*, 1992). However, this variation may not be particularly useful for short-term evolutionary studies, such as population structure and differentiation, as the high mutation rate appears to homogenize conspecific populations, overriding the differentiation caused by gene flow (Fuller, 1991; Arnason & Rand, 1992). This variation is also not suitable for "medium-term" evolutionary studies, such as taxonomic studies at the species, genus or family level because repeated sequences from related species can be drastically different in their DNA sequence. For example, Gjetvaj *et al.* (1992) found that repeated sequences in six different species of scallops had no appreciable sequence similarity with each other and carried little information about species relatedness.

ACKNOWLEDGMENTS

The work reported here was supported by an NSERC research grant and a Network of Excellence (OPEN) grant to E.Z. Amy Ball has helped with the drawing of Figure 1.

LITERATURE CITED

Arnason, E. and D. M. Rand. 1992. Heteroplasmy of short tandem repeats in mitochondrial DNA of Atlantic cod, *Gadus morhua*. Genetics 132:211–220.

Attardi,G. 1985. Animal mitochondrial DNA: an example of genetic economy. International Review of Cytology. 93: 93–145.

Avise,J. C. 1991. Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. Annual Review of Genetics. 25:45–69.

Bentzen, P., W. C. Leggett and G. G. Brown. 1988. Length and restriction site heteroplasmy in the mitochondrial DNA of American shad (*Alosa sapidissima*). Genetics 118:509–518.

Birky, C. W., Jr., T. Maruyama and P. Fuerst. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics 103:513–527.

Birky, C. W.,Jr., P. Fuerst and M. Maruyama. 1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and compositions to nuclear genes. Genetics 121:613–627.

Boyce, T. M., M. E. Zwick and C. F. Aquadro. 1989. Mitochondrial DNA in pine weevils: size, structure and heteroplasmy. Genetics 123:825–836.

Brown, J. R., A. T. Beckenbach and M. J. Smith. 1992. Mitochondrial DNA length variation and heteroplasmy in populations of white sturgeon (*Acipenser transmontanus*). Genetics 132:221–228.

Buroker, N. E., J. R. Brown, T. E. Gilbert, P. J. O'Hara, A. T. Beckenbach, W. K. Thomas and M. J. Smith. 1990. Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics 124:157–163.

Fuller, K. M. 1991. Polymorphism and heteroplasmy of mitochondrial DNA in the scallop *Placopecten magellanicus* (Gmelin). Ph.D. Dissertation, Dalhousie University, Halifax, N.S, Canada.

Fuller, K. M. and E. Zouros. 1993. Dispersed discrete length polymorphism of mitochondrial DNA in the scallop *Placopecten magellanicus* (Gmelin). Current Genetics 23:365–369.

Gjetvaj B., D. I. Cook and E. Zouros. 1992. Repeated sequences and large scale size variation of mitochondrial DNA: A common feature among scallops (Bivalvia: Pectinidae). Molecular Biology and Evolution 9:106–124

Harrison, R. G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends in Ecology and Evolution 4:6–11.

Hauswirth, W. W., M. J. Van De Walle, P. J. Laipis and P. D. Olivo. 1984. Heterogeneous mitochondrial DNA D-loop sequences in bovine tissue. Cell 37:1001–1007.

Hoffmann,R. J., J. L.Boore and W. M. Brown. 1992. A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. Genetics 131:397–412.

Hyman, B. C. and T. M. Slater. 1990. Recent appearance and molecular characterization of mitochondrial DNA deletions within a defined nematode pedigree. Genetics 124: 845–853.

LaRoche, J., M. Snyder, D. I. Cook, K. Fuller and E. Zouros. 1990. Molecular characterization of a repeat element causing large-scale size variation in the mitochondrial DNA of the sea scallop *Placopecten magellanicus*. Molecular Biology and Evolution. 7:45–64.

Meyer, A. 1994. Molecular phylogenetic studies of fishes. *In:* A. R. Beaumont(ed.) Evolution and Genetics of Aquatic Organisms. Chapman and Hall, London (in press).

Moritz, C., T. E. Dowling and W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annual Review of Ecology and Systematics. 18:269–292.

Powers, T. O., E. G. Platzer and B. C. Hyman. 1986. Large mitochondrial genome and mitochondrial size polymorphism in the mosquito parasite *Romanomermis culicivorax*. Current Genetics 11:71–77.

Rand, D. M. and R. G. Harrison. 1986. Mitochondrial DNA transmission genetics in crickets. Genetics 114:955–970.

Rand, D. M. and R. G. Harrison. 1989. Molecular population genetics of mtDNA size variation in crickets. Genetics 121: 551–569.

Solignac, M., J. Genermont, M. Monnerot and J-C. Mounolou. 1984. Genetics of mitochondria in *Drosophila*: mtDNA inheritance in heteroplasmic strains of *D. mauritiana*. Molecular and General Genetics 197:183–188.

Snyder, M., A. R. Fraser, J. LaRoche, K. E. Gartner-Kepkay and E. Zouros. 1987. Atypical mitochondrial DNA from the deep sea scallop *Placopecten magellanicus*. Proceedings of the National Academy of Science. USA 84:7595–7599.

Willis, G. P. 1987. Mitochondrial DNA insertion polymorphism and germ line heteroplasmy in the *Triturus cristatus* complex. Heredity 58:229–238.

Zouros, E., G. H. Pogson, D. I. Cook And M. J. Dadswell. 1992a. Apparent selective neutrality of mitochondrial DNA size variation: A test in the deep sea scallop *Placopecten magellanicus*. Evolution 46:1466–1476.

Zouros, E., K. R. Freeman, A. Oberhauser Ball and G. H. Pogson. 1992b. Direct evidence for extensive mitochondrial DNA inheritance in the marine mussel *Mytilus*. Nature 359:412–414.

Reconstruction of Phylogeny of 11 Species of *Littorina* (Gastropoda: Littorinidae) using Mitochondrial DNA Sequence Data

Elaine Rumbak[1]
David G. Reid
Richard H. Thomas
Department of Zoology
The Natural History Museum
London, SW7 5BD, U.K.

ABSTRACT

A gene phylogeny of eleven *Littorina* and two *Nodilittorina* species was reconstructed using mitochondrial DNA sequence from the small ribosomal RNA gene. The monophyly of the genus *Littorina* is supported as is the inclusion of *L. striata* in this genus. The deep intra-generic branches in *Littorina* are well resolved but the sequence data are insufficiently variable to resolve recent divergences. The phylogenetic trees obtained are consistent with the morphological cladogram of Reid (1990a). These data support in part the biogeographic hypothesis of trans-Arctic migration of *Littorina*.

Key words: Littorina, *Nodilittorina*, phylogeny, 12S rRNA.

INTRODUCTION

Littorina is perhaps the most well-studied of all marine gastropod genera. Nevertheless, only recently has any information been available on the evolutionary relationships within this genus. Using morphological comparisons, sparse palaeontological records, modern distributions and palaeoclimatic data, Golikov & Tzvetkova (1972) discussed in a general way the evolution of *Littorina* from warm-water ancestors in the genus *Nodilittorina*. Cladistic analysis of anatomical characters has been used by Reid (1986, 1989) to attempt a phylogenetic reconstruction for all recognised generic and subgeneric groups in the family Littorinidae. The resulting phylogenetic classification replaced an earlier scheme, which had been based largely on ill-defined features of shell and radula as well as data on penis and sperm (e.g. Rosewater, 1970; Bandel & Kadolsky, 1982). This cladistic analysis defined *Littorina* as a monophyletic group, with sister-taxon *Nodilittorina*. The same method was used to analyse relationships among the 18 species of *Littorina* then recognized (Reid, 1990a); this supported their classification in five subgenera, but failed to resolve completely the branching pattern.

An alternative approach to phylogenetic reconstruction in *Littorina*, independent of morphology, has been the estimation of genetic distance from allozyme data (review by Ward, 1990). Some studies have examined only species from the northern Atlantic (Morris, 1979; Warmoes, 1986; Ward, 1990; Knight & Ward, 1991), and their results are consistent with those from the cladistic analysis of morphology, demonstrating close relationships among members of the *L. saxatilis* complex (*L. saxatilis*, *L. arcana*, *L. nigrolineata*), a greater distance to the pair in the *L. obtusata* complex (*L. obtusata*, *L. mariae*), and a distant relationship of both groups to *L. littorea*. Only two allozyme studies have included examples of *Littorina* species from both the northern Atlantic and the northern Pacific in the same analysis (Boulding, 1990; Zaslavskaya *et al.*, 1992). These agree with the morphological cladogram in uniting members of the subgenus *Neritrema* (*Littorina* species with direct development, i.e. *obtusata* and *saxatilis* complexes, *L. sitkana*, *L. subrotundata* and others not discussed here). However, the branching order of species lower on the cladogram, belonging to the subgenus *Littorina* (including *L. littorea*, *L. squalida*, *L. mandshurica*, *L. brevicula*, *L. plena*, *L. scutulata*) shows discrepancies. The basal species of *Littorina* in Reid's (1989, 1990a) scheme is *L. striata*; this has been included only in the allozyme study of Backlejau & Warmoes (1992), who discussed the taxonomic controversy surrounding the inclusion of this species in *Littorina*, *Nodilittorina* or *Melarhaphe*, although their results supported Reid's classification in *Littorina*.

So far, comparison of DNA sequences has not yet been applied to this problem. Sequence data is now widely used in phylogenetic systematics to determine relationships at all levels (e.g. reviews by Avise, 1989; Hillis & Dixon, 1990; Simon, 1991), although examples among molluscs are rare (e.g. Emberton *et al.*, 1990). Mito-

[1] To whom correspondence should be addressed.

chondrial DNA is particularly useful for investigating a variety of problems in systematics and population genetics, since it is haploid, undergoes little or no recombination and is generally transmitted across generations only by females (Avise *et al.*, 1987). Different mitochondrial genes or parts of genes have been shown to evolve at a wide range of rates; rapidly evolving DNA sequences can be used to study recently diverged taxa, whereas conserved regions are suitable for distantly related taxa (Simon, 1991). The advent of modern molecular techniques has made it possible to sample DNA sequence variation from selected regions of these genomes quickly and from large numbers of individuals.

In this paper we use data obtained from direct sequencing of amplified mitochondrial DNA, using universal primers for part of the small ribosomal RNA gene (Kocher *et al.*, 1989). The sequence data are used to construct a species-level gene phylogeny of the eleven northern Atlantic and eastern Pacific *Littorina* species so far available to us. This will provide a test, independent of both morphology and allozymes, for the cladogram of Reid (1989, 1990a). By including *L. striata* together with two members of the sister-taxon *Nodilittorina*, the disputed affinities of this species are investigated.

MATERIALS AND METHODS

Sample collection and DNA extraction: The different species, numbers of individuals analysed, and collection sites of *Littorina* used in this study are listed in Table 1. After collection the animals were kept alive at 4°C for not more than a week. Before dissection the animals were anaesthetised in 7.5% magnesium chloride (w/v) solution for 1 hour. Sex was determined by presence of penis or pallial oviduct, and any individuals parasitised by trematodes were discarded. The operculum and rectum, with faecal pellets, were removed. For the larger species (>10 mm shell length) only the digestive gland was removed for DNA extraction, whereas for the smaller species the whole animal was processed. For DNA extraction a modification of the guanidinium thiocynate method of Pitcher *et al.* (1989) was used. The coarsely chopped animal was placed in equal volumes of Tris-EDTA (0.1M, pH 8.0) and GE reagent (5M guanidinium thiocynate, 1M EDTA), a half volume of glass beads (Glasperlen, Braun) added, and the sample shaken (2 min, 4°C, 2,000 rpm) in a bead beater (Mikro-Dismembrator, Braun). The beads and debris were removed by centrifugation (13,000 × g, 15 min, 4°C). The supernatant was collected and half a volume of ammonium acetate (7.5 M) added, before treatment with phenol and chloroform to remove proteins and pigments. The DNA in the sample was precipitated with ethanol, washed with 70% ethanol, dried, and resuspended in sterile distilled water. The samples were then stored at 4°C, or for long-term storage, −70°C.

Polymerase Chain Reaction (PCR) and DNA sequencing: Amplification of part of the mitochondrial small ribosomal RNA gene (12S rRNA) was carried out using the universal 12Sa and 12Sb primers (Kocher *et al.*, 1989). Each PCR was performed in a 100 μl volume consisting of 67 mM Tris-HCL (pH8.8), 2 mM $MgCl_2$, 0.05% Tween–20, 100 μg/ml bovine serine albumin, 40 pmoles of each primer, 100 μM of each dNTP, 2.5 units Taq Polymerase (Perkin-Elmer/Cetus), and 10–1000 ng template DNA. The cycling parameters for amplification were an initial 5 min denaturation at 94°C, and then 30 cycles of 45 sec at 94°C, 1.5 min at 55°C, and 2 min at 72°C. Precipitated PCR products were then directly sequenced by the rapid thermal cycling technique described in Embley (1991). DNA sequence was determined for both strands. The DNA sequences have GenBank accession numbers u@5862–u@5874-v.

Sequence analysis and gene phylogeny reconstruction: Sequences were aligned with the aid of the program CLUSTALV (Higgins *et al.*, 1992), with minor adjustments made by eye. Results from a range of analytical methods were compared (review by Swofford and Olsen 1990). Analyses employing maximum parsimony criteria were carried out with PAUP, version 3.0s (Swofford, 1990). Bootstrapped distance-based analyses were done with CLUSTALV (Higgins *et al.*, 1992), which employs Kimura's (1980) 2-parameter model for the calculation of corrected pairwise distances and the neighbour-joining algorithm of Saitou and Nei (1987) for the construction of trees. The maximum likelihood method (Felsenstein, 1981) was carried out with PHYLIP, version 3.4 (Felsenstein, 1991). For further details of methodology consult legends to figures 2 and 3.

RESULTS

Sequences of the 12S rRNA gene fragment (not including the primer sequence) from eleven *Littorina* species and two *Nodilittorina* species are shown aligned in Figure 1. No within-species variation was found for any of the species used (for sample size see Table 1) except for *L. saxatilis*. Both the South African and Welsh *L. saxatilis* populations had two non-identical single base pair differences, as well as a sequence in common with Isle of Wight and Venice populations. Only the sequence in common is presented here (Figure 1). Alignment of the sequences was not difficult because this gene was found to be fairly well conserved. There are 73 variable positions (including deletions and insertions) among the 374 sites aligned.

Phenetic analysis of the sequence data for each pair of taxa is presented in Table 2. The total number of transversions and transitions, as well as the number of transversions relative to transitions, increase from left to right in the table (i.e. *L. saxatilis* to *N. trochoides*). Likewise, the genetic distance (calculated from Kimura's (1980) 2-parameter estimator, which allows transitions and transversions to occur at different rates) increases from top to bottom in the species matrix. Since the order of the species in Table 2 is based on the morphological cladogram of Reid (1990a), these trends reflect the increasingly distant relationships between species. These data show the expected transition bias (Brown *et al.*,

```
N. radiata      TCTTAGGC-A TAAATAAATT TAAATATTTA CTAGAGTACT ACGAATAAAA  50
N. trochoides  ......GT-A .....TG..- .GGC...... .C........ .....TA.A.  50
L. striata     .TC...AC-G .....C..C. CCG---.... .......... .....TT.T.  50
L. keenae      .......C-. .....CG... .A.--..C.. ..G....... .....TC...  50
L. scutulata   ........-. .......... .C.--..... .......... .....C....  50
L. plena       ........-. .......... .C.--..... .......... ..........  50
L. littorea    ........-. .......... .T.--..... .......... ..........  50
L. subrotundata ........-. .......... ...--..... .......... ..........  50
L. obtusata    ........-. .......... ...--..... .......... ..........  50
L. mariae      ........-. .......... ...--..... .......... ..........  50
L. nigrolineata ........C. .......... ...--..... .......... ..........  50
L. arcana      ........-. .......... ...--..... .......... ..........  50
L. saxatilis   ........-. .......... ...--..... .......... ..........  50

N. radiata     CTATTTAAAA CTCAAAGAGC TTGGCGGTGC TTTAGACTTC TCAGGGGAAC 100
N. trochoides  CC........ .......... .......... T........C .......... 100
L. striata     A-........ .......... .......... T........T .......... 100
L. keenae      .-........ .......... .......... T..------- .......... 100
L. scutulata   .-........ .......... .......... C..------- ...A...... 100
L. plena       .-........ .......... .......... .......... .......... 100
L. littorea    .-........ .......... .......... .......... .......... 100
L. subrotundata .-........ .......... .......... .......... .......... 100
L. obtusata    .-........ .......... .......... .......... .......... 100
L. mariae      .-........ .......... .......... .......... .......... 100
L. nigrolineata .-........ .......... .......... .......... .......... 100
L. arcana      .-........ .......... .......... .......... .......... 100
L. saxatilis   .-........ .......... .......... .......... .......... 100

N. radiata     CTGTCTCGTA ATCGACAGTC CACGAATCAA CCTTACCTTC TTTCGCG-AT 150
N. trochoides  .......G.. .......G.. ........T. ...C..TT.C C..T.CGT.. 150
L. striata     .....C.G.. .......A.. ......A... .......TCT ...C.CAC-. 150
L. keenae      .......A.. .......... .C........ .......T.. C....CGA.. 150
L. scutulata   .......... .......... ......-A.. .......C.. .....T..G- 150
L. plena       .......G.. .......... ......-A.. .......... .........- 150
L. littorea    .......... .......... ......-... .......... .....C...- 150
L. subrotundata .......... .......... ......-... .......... .....C...- 150
L. obtusata    .......... .......... ......-... .......... .........- 150
L. mariae      .......... .......... ......-... .......... .........- 150
L. nigrolineata .......... .......... ......-... .......... ........G- 150
L. arcana      .......... .......... ......-... .......... ........G- 150
L. saxatilis   .......... .......... ......-... .......... ........G- 150

N. radiata     CAGTATGTAT ACCGTCGTCG TCAGGTAACT TTTAAAAATA TAGAAGTTAG 200
N. trochoides  .......... .......... T......... ...G...... ..GA...... 200
L. striata     .......... .......... C......... .C........ C.GC...... 200
L. keenae      .......... .......... T......... .......G.. ..AA...... 200
L. scutulata   .......... .......... .......... ........C. .......... 200
L. plena       .......... .......... T......... ...T...... .......... 200
L. littorea    .......... .......... .......... .......... ....T..... 200
L. subrotundata .......... .......... .......... .......... .......... 200
L. obtusata    .......... .......... .......... .......... .......... 200
L. mariae      .......... .......... T......... .......... .......... 200
L. nigrolineata .......... .......... .......... .C........ .... ..... 200
L. arcana      .......... .......... .......... .C........ .......... 200
L. saxatilis   .......... .......... .......... .C........ .......... 200

N. radiata     CAACAAAACT ATTAAGTTTT TACGTCAGAT CAAGGTGCAG CTTATAAGAA 250
N. trochoides  .A.G....CC T......... .......... .......... .C...GGG.A 250
L. striata     .A.AG..GC. ...-..C..C .......... .......... .T....TAGA 250
L. keenae      .TT....CCC T.--...... .......... .......... ......G..A 250
L. scutulata   .C......TC .....A.... .......... .......... .........G 250
L. plena       ....G..... .......... .......... .......... .......... 250
L. littorea    ....G...C. ...T...... C......... .......... .......... 250
L. subrotundata .C........ .......... .......... .......... .......... 250
L. obtusata    .......... .......... .......... .......... .......... 250
L. mariae      .......... .......... .......... .......... .......... 250
L. nigrolineata ..T.....C. .......... .......... .......... .......... 250
L. arcana      .......... .......... .......... .......... .......... 250
L. saxatilis   .......... .......... .......... .......... .......... 250

N. radiata     GGAGAAGATG GGTTACAATT AAT-ATTTAT AACTACGGAT ATAAGTTTTA 300
N. trochoides  A...CG.... .......... ..T-...... .......... .TCAATCTTG 300
L. striata     ....CG.... .......... .TTT...C.. ..TA...... .TTAACTTT. 300
L. keenae      ..G....... .......... ..A-.C.... .......... .T.G..TAT. 300
L. scutulata   .......... .......... ...-.C.... ..T....... .A.A....G. 300
L. plena       .......... .......... ...-...... .......... .G......GG 300
L. littorea    .......... .......... ...-...... ..T....... .....T..A. 300
L. subrotundata .......... .......... ..T-...... .......A.. .......... 300
L. obtusata    .......... .......... ...-...... .......... ........G. 300
L. mariae      .......... .......... ...-...... .......... ........G. 300
L. nigrolineata .......... .......... ...-...... .......... .......... 300
L. arcana      .......... .......... ...-...... .......... .......... 300
L. saxatilis   .......... .......... ...-...... .......... .......... 300

N. radiata     A-ATGACTTA TATGAAGGCG GACTTGAAAG TATGAATTAG TATAGAAATT 350
N. trochoides  .-.TGGTTTA .A......C. .......... ..TG.A.... ..TAT....T 350
L. striata     .C.--.TT-T .T......C. .......... ..AACT.... C.GTG....A 350
L. keenae      .T.AA..CAT ........A. .......... .....GC... ..GCT....G 350
L. scutulata   .--....T.C ..-..... .......... .......... ..ACT....A 350
L. plena       .--..... ..-A..... .......... .......... ....C..... 350
L. littorea    .--..... ..-..... .......... .......... .......... 350
L. subrotundata .--..... ..-..... .......... .......... .......... 350
L. obtusata    .--...T... ..-..... .......... .......... C...T..... 350
L. mariae      .--..... ..-..... .......... .......... C...T..... 350
L. nigrolineata .--..... ..-..... .......... .......... .......... 350
L. arcana      .--..... ..-..... .......... .......... .......... 350
L. saxatilis   .--..... ..-..... .......... .......... ....G..... 350

N. radiata     TTGTGAATCT AGCTCTGAAG ACTG                          374
N. trochoides  T.G.....TT ........A. CG..                          374
L. striata     T.T.....CA ........A. ....                          374
L. keenae      C......... ........A. ....                          374
L. scutulata   ........C. ........G. ....                          374
L. plena       ........C. .......... ....                          374
L. littorea    .......... .......... ....                          374
L. subrotundata ........C. .......... ....                          374
L. obtusata    .......... .......... ....                          374
L. mariae      .......... .......... ....                          374
L. nigrolineata .......... .......... ....                          374
L. arcana      .......... .......... ....                          374
L. saxatilis   .......... .......... ....                          374
```

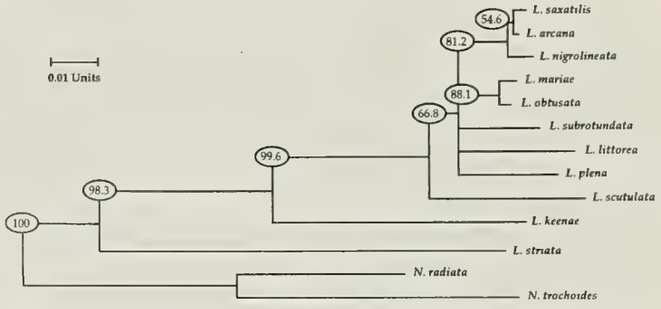


Figure 2. Bootstrapped neighbour-joining tree constructed with pairwise distances calculated using Kimura's (1980) 2-parameter correction for multiple substitutions. Branches are drawn proportional to their length. Produced with CLUSTALV (Higgins *et al.*, 1992). Sites at which any species has a gap have been included in the analysis. Numbers on nodes are the percentage of replicates in which the taxa to the right of the node occur together. This is a bootstrap 50% majority-rule consensus tree based on 1000 replicates. The tree is shown with the *Nodilittorina* species as an outgroup to root the tree. Arbitrarily groups with less than 50% support have been collapsed to polytomies.

1982; Moritz *et al.*, 1987) seen in other metazoan mitochondrial sequences, and judging from the relatively low ratio of transversions to transitions these sequences are not yet saturated, even for the deepest divergences. In all pairwise comparisons the number of transitions exceed the number of transversions (Table 2). In general, once transitions have become saturated, the transversions continue to accumulate approximately linearly with time well past the point at which transitions have reached their asymptotic value (Miyamoto & Boyle, 1989).

Phylogenetic trees constructed using neighbour-joining and maximum parsimony are shown in Figures 2 and 3 respectively. These trees are consistent with each other, though the neighbour-joining tree shows slightly more resolution of *L. scutulata* and of the *L. saxatilis* complex. The maximum likelihood method tree (not shown) had the same topology as the most parsimonious tree. In the parsimony analysis a single most parsimonious tree was obtained of length 185 steps. By permitting the tree length to increase by 1 step, and taking a strict consensus of the resulting trees, it was found that only four nodes remained in the ingroup (supporting *L. striata, L. keenae,* the remaining species with the *L. saxatilis* complex). At 187 steps the *L. saxatilis* complex collapsed; the next branch to collapse was that separating *L. keenae* and *L. striata* at 190 steps.

←

Figure 1. Aligned DNA sequences of a fragment of the mitochondrial 12S rRNA gene from 13 species of littorinid gastropods. Dots indicate that the sequence is identical to *N. radiata*, dashes indicate deletions in one or more sequences relative to other sequences shown.

Table 1. Localities and sample sizes of *Nodilittorina* and *Littorina* used in this study.

Species	Locality	Sample size
N. trochoides (Gray, 1839)	Cape d'Aguilar, Hong Kong	2
N. radiata (Eydoux & Souleyet, 1852)	Cape d'Aguilar, Hong Kong	2
L. striata King & Broderip, 1832	El Golfo, Lanzarote, Canary Is.	4
L. keenae Rosewater, 1978	Pacific Grove, California, USA	4
L. scutulata Gould, 1849	Pacific Grove, California, USA	3
L. plena Gould, 1849	Candlestick Park, San Francisco, California, USA	3
L. littorea (Linnaeus, 1758)	St. Lawrence, Isle of Wight, UK	6
	West Angle Bay, nr Pembroke, Wales, UK	3
L. subrotundata (Carpenter, 1864)	Charleston, Oregon, USA	3
L. obtusata (Linnaeus, 1758)	Pembroke Dock, Wales, UK	3
L. mariae Sacchi & Rastelli, 1966	West Angle Bay, nr Pembroke, Wales, UK	3
L. nigrolineata Gray, 1839	West Angle Bay, nr Pembroke, Wales, UK	3
L. arcana Hannaford Ellis, 1978	St. Govan's Head, nr Pembroke, Wales, UK	3
L. saxatilis (Olivi, 1792)	St. Govan's Head, nr Pembroke, Wales, UK	2
	Alberoni, Venice, Italy	2
	Langebaan Lagoon, South Africa	4
	St. Lawrence, Isle of Wight, UK	2

DISCUSSION

The branches nearer the base of both trees (Figures 2 and 3) are well resolved and are consistent with Reid's (1989, 1990a) concept of the monophyletic genus *Littorina* with *L. striata* as its basal member. Strictly, our data provide only a partial test of the inclusion of *L. striata* in *Littorina*; clearly it does not cluster between the two *Nodilittorina* species, but the hypothesis that it could be a basal member of *Nodilittorina* cannot be falsified without recourse to more distant outgroups. However, it may be noted that the calculated genetic distances (Table 2) show that in 6 of the 10 possible comparisons with other *Littorina* species, *L. striata* is closer to the *Littorina* than to either of the *Nodilittorina* species. This is reflected in the branch lengths of Figure 2. In addition, midpoint rooting of the maximum parsimony tree placed the tree root between the *Nodilittorina* species and *L. striata*. Previously there has been some debate about the classification of *L. striata*, with most recent authors placing it in *Nodilittorina* on morphological grounds (Rosewater, 1981; Bandel & Kadolsky, 1982). However, in their analysis of allozyme data Backeljau & Warmoes (1992) also favoured inclusion in *Littorina*. The terminal groupings of *L. saxatilis*, *L. arcana* and *L. nigrolineata* (the *L. saxatilis* complex) and *L. obtusata* and *L. mariae* (the *L. obtusata* complex) are also supported, agreeing with the results of both morphological (Reid, 1989; 1990a) and allozyme studies (Warmoes, 1986; Ward, 1990; Knight & Ward, 1991; Zaslavskaya *et al.*, 1992).

At the more intermediate levels, however, neither tree is well resolved. The maximum parsimony tree (Figure 3) is entirely consistent with the morphological cladogram of Reid (1990a; Figure 4), but does not resolve relationships between the members of the subgenera *Littorina* (*L. scutulata*, *L. plena*, *L. littorea*) and *Neritrema* (*L. saxatilis* complex, *L. obtusata* complex, *L. subrotundata*). The neighbour-joining tree shows slightly more

Table 2. The number of transversion differences followed by the number of transition differences (above the diagonal) for the 374 sites of the 12S rRNA gene fragment, and the pairwise genetic distance calculated from Kimura's (1980) 2-parameter estimator (below the diagonal). The first three letters of species names are used as abbreviations above columns.

	sax	arc	nig	mar	obt	sub	lit	ple	scu	kee	str	rad	tro
L. saxatilis	—	1/0	2/1	1/5	1/5	2/6	3/8	3/10	2/16	11/25	14/41	25/34	27/41
L. arcana	0.003	—	1/1	0/6	0/6	1/6	2/8	2/10	1/17	10/26	15/41	26/34	26/42
L. nigrolineata	0.008	0.006	—	1/7	1/7	2/7	3/7	3/11	2/18	9/25	16/40	27/33	27/41
L. mariae	0.017	0.017	0.022	—	0/2	1/8	2/10	2/8	1/17	10/23	15/24	26/32	26/39
L. obtusata	0.017	0.017	0.022	0.006	—	1/8	2/10	2/10	1/17	10/25	15/40	26/34	26/39
L. subrotundata	0.022	0.020	0.025	0.025	0.025	—	3/8	3/10	2/17	11/26	14/41	25/31	25/41
L. littorea	0.031	0.028	0.028	0.034	0.034	0.031	—	4/12	3/21	11/26	16/40	28/32	28/40
L. plena	0.037	0.034	0.040	0.028	0.034	0.037	0.046	—	1/19	12/28	17/41	28/32	28/39
L. scutulata	0.052	0.052	0.058	0.052	0.052	0.055	0.070	0.058	—	11/29	16/42	27/36	27/43
L. keenae	0.108	0.108	0.102	0.098	0.105	0.111	0.112	0.121	0.121	—	20/41	28/30	24/35
L. striata	0.177	0.180	0.180	0.184	0.176	0.177	0.180	0.187	0.187	0.195	—	25/33	22/43
N. radiata	0.185	0.189	0.188	0.181	0.189	0.174	0.188	0.188	0.200	0.180	0.183	—	10/23
N. trochoides	0.219	0.219	0.219	0.207	0.207	0.212	0.219	0.215	0.227	0.184	0.210	0.096	—

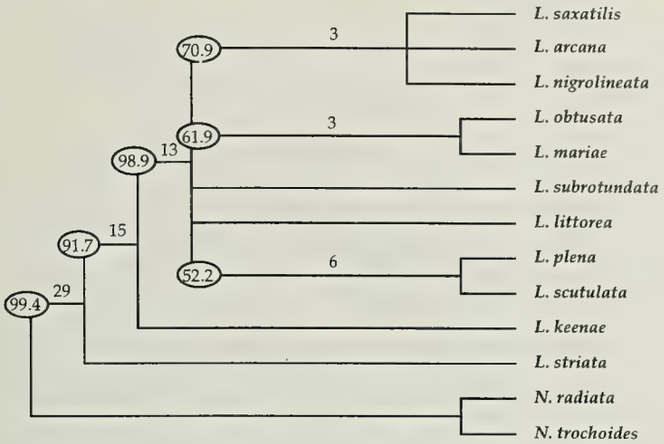


Figure 3. Bootstrapped maximum parsimony tree based on analysis of 12S rRNA gene sequences produced with PAUP (Swofford, 1990) using a branch-and-bound search. Options used were 'furthest' addition sequence; ACCTRAN; all minimal length trees saved (MULPARS); zero-length branches collapsed; gaps treated as missing (treatment of gaps as a fifth character has no effect on the topology and has only a slight effect on bootstrap values). Numbers on branches are the number of inferred synapomorphies. The numbers on the nodes are the percentage of replicates in which the taxa originating from there occur together. Bootstrap parameters are as described for Figure 2. The tree is shown with the *Nodilittorina* species as an outgroup to root the tree. There are 158 character states distributed over 57 informative characters. The tree is 185 steps long.

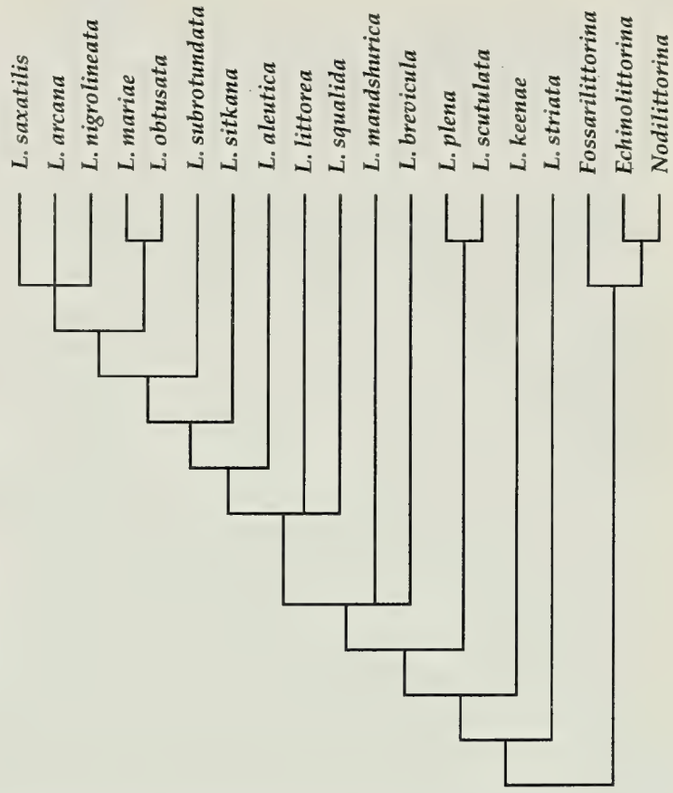


Figure 4. Morphological cladogram of 16 *Littorina* species, redrawn from Reid (1990a). The following modifications were made: exclusion of *Mainwaringia rhizophila* (now not considered a member of this clade, unpublished data); exclusion of *L. kurila* (a synonym of *L. subrotundata)* and *L. neglecta* (a probable synonym of *L. saxatilis).*

resolution, as a result of the contribution of autapomorphies to the pairwise distances among taxa for which there are few or no sites that are informative for parsimony analysis. This tree (Figure 2) separates *L. scutulata* in the same position as in the morphological cladogram, but does not place *L. plena* as its expected sister-species (Reid, 1990a). However, in view of the relatively low bootstrap values at the node separating *L. scutulata*, little confidence can be placed in this result. Disagreements as to the position of *L. scutulata* and its probable sister-species (Boulding, 1990; Reid, 1990a) are therefore not yet resolved. Other disagreements between the morphological cladogram and allozyme-based trees (Zaslavskaya *et al.*, 1992) involve species from the northwestern Pacific that have yet to be included in our study.

The morphological cladogram of Reid (1990a) has been used as evidence of the origin of the Atlantic species of *Neritrema* from a Pacific ancestor, an example of the migration of marine fauna from the northern Pacific into the northern Atlantic following the opening of the Bering Strait during the Upper Pliocene (Reid, 1990b). Although the present results do not resolve the relationships of *Neritrema*, there is some support for this biogeographic hypothesis, in that the Atlantic *Neritrema* species show the lowest genetic distances among themselves (Table 2, Figure 2).

The distance measures given in Table 2 show a close relationship among species classified in the subgenera *Littorina* and *Neritrema*, which is surprising in view of their rather marked morphological differences (Reid, 1989; 1990a). Both *L. striata* and *L. keenae* are relatively distant from these and from each other, supporting their classification in separate subgenera (Reid, 1989). The two species of *Nodilittorina* used as the outgroup in this analysis, *N. radiata* and *N. trochoides*, are also relatively distant from each other; their shells are dissimilar, but anatomically they show few differences and are presently placed in the same subgenus, *Nodilittorina*. There has not yet been a species-level analysis of phylogenetic relationships in *Nodilittorina*.

The inference models used in this study to analyse the sequence data do not assume a 'molecular clock' (Swofford & Olsen, 1990), and it is uncertain if the same topology would result if a clock were assumed. If, however, the 'molecular clock' hypothesis, that genetic distance measures are proportional to the time of phylogenetic divergence, is tentatively assumed, one could make the following crude estimates. For *Littorina*, two nodes on the cladogram can be dated: the separation of *L. littorea* from its Pacific sister-species *L. squalida*, and the separation of the Atlantic *Neritrema* species from the sister-taxon of this clade in the Pacific. The earliest possible date for both of these separations is the opening of the Bering Strait, dated at 3.5–4.0 million years (My)

ago (Hopkins, 1967; Reid, 1990b; Vermeij, 1991). The latest date is taken as the time of the onset of widespread glaciation about 2.4 My (Shackleton *et al.*, 1984; Loubere, 1988) when the trans-Arctic migration route may have been cut by climatic cooling. *Littorina squalida* was not included in this study, but the sister-taxon of the Atlantic *Neritrema* species is believed to be *L. subrotundata* (Reid, 1990a). The average distance between *L. subrotundata* and the five Atlantic *Neritrema* species is 0.023 (Table 2), a divergence of 0.00575 per My (accepting the older estimate of the age of the Bering Strait). This approximation permits tenuous estimates of the ages of other nodes on the tree: 8 My for the separation of *L. plena* plus *L. scutulata* (assuming these are sister-taxa) from more recent species; 19 My for *L. keenae*; 32 My for *L. striata* and therefore a minimum age of the clade as a whole. These ages are probably too young and would have become even more so had the latest separation date been used in the calibration. The fossil record of *Littorina* is very poor, and the older fossils, dating to the Upper Palaeocene, cannot be assigned to the genus with any confidence (Reid, 1989; 1990b). The oldest certain member of the genus is *L. sookensis* from the lower Miocene of Vancouver Island (approximately 22 My), which is probably close to the modern *L. keenae*. The Recent species *L. squalida* (sister-species of *L. littorea*) has been recorded from the Middle Miocene of Kamchatka (approximately 15 My). It should be stressed that this means of estimation of ages from molecular data gives only very approximate results.

This study has shown that the 12S rRNA gene is insufficiently variable to resolve the relationships among recently diverged species such as members of the subgenera *Littorina* and *Neritrema*. However, it has permitted well-supported resolution of the deeper branches of the phylogeny. This is part of an ongoing study and we now intend to sequence more variable portions of the mitochondrial genome to resolve relationships more fully, including those at the intraspecific level. We propose to examine as many as possible of the 21 species presently classified as *Littorina*, in the attempt to derive a well-resolved phylogeny for the genus. This will provide an evolutionary framework for the large amount of ecological and physiological data already available for the genus, and thus permit tests of adaptational hypotheses (Reid, 1989; 1990a). It will also test hypotheses about speciation and biogeography of marine invertebrates in temperate latitudes for which this well-studied genus has been used as an example (Golikov & Tzvetkova, 1972; Reid, 1990b; Vermeij, 1991).

ACKNOWLEDGEMENTS

This work was supported by NERC grant GR3/7854. We would like to thank Stephen A. Ridgway and James T. Carlton for providing some of the samples used. Assistance in the preparation of figures was provided by Gail Altschuler.

LITERATURE CITED

Avise, J. C. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution 43: 1192–1208.

Avise, J. C., J. Arnold, R. Martin Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics 18:489–522.

Backeljau, T. and T. Warmoes. 1992. The phylogenetic relationships of ten Atlantic littorinids assessed by allozyme electrophoresis. *In:* Grahame, J., P. J. Mill and D. G. Reid (eds.). Proceedings of the third International Symposium on Littorinid Biology pp.9–24. London: Malacological Society of London.

Bandel, K. and D. Kadolsky. 1982. Western Atlantic species of *Nodilittorina* (Gastropoda: Prosobranchia): comparative morphology and its functional, ecological, phylogenetic and taxanomic implications. Veliger 25:1–42.

Boulding, E. G. 1990. Systematics, ecology and ecological genetics of some northern Pacific *Littorina*. PhD thesis, University of Washington, Seattle.

Brown, W. M., E. M. Prager, A. Wang and A. C. Wilson. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. Journal of Molecular Evolution 18:225–239.

Emberton, K. C., G. S. Kuncio, G. M. Davies, S. M. Phillips, K. M. Monderewicz, and H. G. Yuan. 1990. Comparison of recent classifications of stylommatophoran land-snail families, and evaluation of large-ribosomal-RNA sequencing for their phylogenetics. Malacologia 31:327–352.

Embley, T. M. 1991. The linear PCR reaction: a simple and robust method for sequencing amplified rRNA genes. Letters in Applied Microbiology 13:171–174.

Felsenstein, J. 1981. Evolutionary trees from DNA Sequences: A Maximum Likelihood Approach. Journal of Molecular Evolution 17:368–376.

Felsenstein, J. 1991. PHYLIP (phylogeny inference package) version 3.4. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Golikov, A. N., and N. L. Tzvetkova. 1972. The ecological principle of evolutionary reconstruction as illustrated by marine animals. Marine Biology 14:1–9.

Higgins, D. G., A. J. Bleasby and R. Fuchs. 1992. CLUSTAL-V-Improved software for multiple sequence alignment. Computer Applications in the Biosciences 8:189–191.

Hillis, D. M. and M. T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. The Quarterly Review of Biology 66:411–453.

Hopkins, D. M. 1967. The Cenozoic history of Beringia- a synthesis. *In:* Hopkins, D. M. (ed.) The Bering Land Bridge : pp. 451–484. Stanford, California; Stanford University Press.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111–120.

Knight, A. J. and R. D. Ward. 1991. The genetic relationships of three taxa in the *Littorina saxatilis* species complex (Prosobranchia: Littorinidae). Journal of Molluscan Studies 57:81–91.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca and A. C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: am-

plification and sequencing with conserved primers. Proceedings of the National Academy of Sciences, USA 86: 6196–6200.

Loubere, P. 1988. Gradual late Pliocene onset of glaciation; a deep-sea record from the northeast Atlantic. Palaeogeography, Palaeoclimatology and Palaeoecology 63:327–334.

Miyamoto, M. M., and S. M. Boyle. 1989. The potential importance of mitochondrial DNA sequence data to eutherian mammal phylogeny. *In:* B. Fernholm, K. Bremer and H. Jornvall (eds). The hierarchy of life. Elsevier Science, Amsterdam.

Moritz, C., T. E. Dowling and W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annual Review of Ecology and Systematics 18:269–292.

Morris, S. R. 1979. Genetic variation in the genus *Littorina*. Unpublished thesis, University of Wales.

Pitcher, D., N. A. Saunders and R. J. Owen. 1989. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Letters in Applied Microbiology 8:151–156.

Reid, D. G. 1986. The Littorinid molluscs of Mangrove forests in the Indo-Pacific region. The genus *Littoraria*. British Museum (Natural History), London, 228p.

Reid, D. G. 1989. The comparative morphology, phylogeny and evolution of the gastropod family Littorinidae. Philosophical Transactions of the Royal Society of London, Series B 324:1–110.

Reid, D. G. 1990a. A cladistic phylogeny of the genus *Littorina* (Gastropoda): implications for evolution of reproductive strategies and for classification. Hydrobiologia 193: 1–19.

Reid, D. G. 1990b. Trans-Arctic migration and speciation induced by climate change: the biogeography of *Littorina* (Mollusca: Gastropoda). Bulletin of Marine Science 47:35–49.

Rosewater, J. 1970. The family Littorinidae in the Indo-Pacific. I. The subfamily Littorininae. Indo-Pacific Mollusca 2:417–506.

Rosewater, J. 1981. The family Littorinidae in tropical West Africa. Atlantide Report, Copenhagen. 13:7–48.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406–425.

Shackleton, N. J., J. Backman, H. Zimmerman, D. V. Kent, M. A. Hall, D. G. Roberts, D. Schnitker, J. G. Baldauf, A. Desprairies, R. Homrighausen, P. Huddlestun, J. B. Keene, A. J. Kaltenback, K. A. O. Krumsiek, A. C. Morton, J. W. Murray and J. Westberg-Smith. 1984. Oxygen isotope calibration of the onset of ice-rafting and history of glaciation in the North Atlantic region. Nature 307:620–623.

Simon, C. 1991. Molecular systematics at the species boundary: exploiting conserved and variable regions of the mitochondrial genome of animals via direct sequencing from amplified DNA. *In:* Hewitt, G. M., A. W. B. Johnston and J. P. W. Young (eds). Molecular Techniques in Taxonomy. Springer, Berlin. p.33–71.

Swofford, D. L. 1990. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0, Illinois Natural History Survey, Champaign, Illinois.

Swofford, D. L. and G. J. Olsen. 1990. Phylogeny reconstruction. *In:* Hillis, D. M. and C. Moritz (eds). Molecular Systematics. Sinauer, Sunderland, Mass. p.411–501.

Vermeij, G. J. 1991. Anatomy of an invasion: the trans-Arctic interchange. Paleobiology 17:281–307.

Ward, R. D. 1990. Biochemical genetic variation in the genus *Littorina*. Hydrobiologia 193: 53–69.

Warmoes, T. 1986. Een inleidende systematische en taxonomische studie van het genus *Littorina* (Gastropoda: Prosobranchia). Licentiaatsthesis, Universitaire Instelling Antwerpen, p.128

Zaslavskaya, N. I., S. O. Sergievsky and A. N. Tatarenkov. 1992. Allozyme similarity of Atlantic and Pacific species of *Littorina* (Gastropoda: Littorinidae). Journal of Molluscan Studies 58:377–84.

Small Ribosomal Subunit RNA and the Phylogeny of Mollusca

Birgitta Winnepenninckx
Departement Biochemie
Universiteit Antwerpen (U.I.A.)
Universiteitsplein 1,
B–2610 Antwerpen, Belgium

Thierry Backeljau
Afdeling Malacologie
Koninklijk Belgisch Instituut voor
Natuurwetenschappen
Vautierstraat 29
B–1040 Brussel, Belgium

Rupert De Wachter[1]
Departement Biochemie
Universiteit Antwerpen (U.I.A.)
Universiteitsplein 1,
B–2610 Antwerpen, Belgium

ABSTRACT

We determined the complete sequence of the small ribosomal subunit RNA of the pulmonate snail *Onchidella celtica*. This sequence and the one recently determined for the chiton *Acanthopleura japonica* were added to an alignment of 25 18S rRNA sequences of Metazoa, including three other Mollusca. The data set was used to assess certain aspects of molluscan phylogeny by distance matrix and character state methods. The trees obtained were tested for effects of random and systematic errors. The results of our analyses support: (a) molluscan monophyly; (b) gastropod monophyly; (c) bivalve monophyly; (d) a sister group relationship of Gastropoda and Polyplacophora. The position of the phylum among other Metazoa remains uncertain due to a lack of representatives of many invertebrate phyla in our data set. Most of our results are congruent with existing hypotheses.

Key Words: 18S rRNA, phylogeny, Metazoa, Gastropoda, Bivalvia, Polyplacophora, *Onchidella celtica*.

INTRODUCTION

Historical Background

Many aspects of molluscan phylogeny are still uncertain. The huge phenotypic diversity within the phylum obscures the evolutionary relationships between the larger molluscan taxa (*e.g.* von Ihering, 1876; Milburn, 1960; von Salvini-Plawen, 1969, 1972, 1985, 1990a,b; Stasek, 1972; Götting, 1980; Wingstrand, 1985; Scheltema, 1988; Brusca & Brusca, 1990). Nevertheless, it is generally accepted that the "shell- bearing molluscs" (Conchifera, *i.e.* Cephalopoda, Scaphopoda, Bivalvia, Gastropoda and Monoplacophora) are monophyletic with Polyplacophora as sister group (*e.g.* von Salvini-Plawen, 1969, 1985, 1990a; Stasek, 1972; Götting, 1980; Wingstrand, 1985; Scheltema, 1988; Brusca & Brusca, 1990). Indeed, the loss of spicules and the presence of three mantle margin folds, an univalve shell consisting basically of three layers, jaws, a head with cerebrally innervated appendages, a nervous system differentiated in axons and ganglia, a crystalline style and statocysts are considered to be synapomorphies uniting the five conchiferan classes (*e.g.* Götting, 1980; von Salvini-Plawen, 1985; Wingstrand, 1985; Brusca & Brusca, 1990). However, different interpretations exist about the phylogenetic relationships within this subphylum. Milburn (1960) suggested three conchiferan clades: Monoplacophora, Cephalopoda and a Bivalvia-Gastropoda-Scaphopoda clade. The branching pattern of these three groups and the topology of the Bivalvia-Gastropoda-Scaphopoda clade, remains unresolved. Götting (1980) proposed a Bivalvia-Scaphopoda sister relationship and relied on the shell structure and form of the larval shell to conclude that Gastropoda and Monoplacophora are sister groups. Cephalopoda is then a sister group to the other four conchiferan classes. However, Wingstrand (1985), Brusca and Brusca (1990) and von Salvini-Plawen (1985, 1990a) considered "Monoplacophora" (*i.e.* class Tergomya *sensu* Peel, 1991 or Tryblidiida *sensu* Wingstrand, 1985; see Peel, 1991 for a discussion) as a sister group to the four other conchiferan classes, which in turn consist of a Bivalvia-Scaphopoda clade and a Gastropoda-Cephalopoda clade. The former is characterized by the presence of a mantle surrounding the entire body, reduction of the head and a laterally compressed form. The latter is determined by the presence of a well developed head, dorsoventral elongation, dorsal concentration of the viscera and shell coiling (Brusca & Brusca, 1990). Except for the position of the "Monoplacophora", this view agrees well with the division of the Conchifera into the clades Diasoma (classes Bivalvia, Scaphopoda and the fossil Rostroconchia) and Cyrtosoma (classes "Monoplacophora", Gastropoda and Cephalopoda), which is widely accepted among paleontologists (Runnegar & Pojeta, 1974, 1985; Pojeta, 1980; Steiner, 1992). Yet, Peel (1991) recently suggested that Cyrtosoma and Diasoma are both polyphyletic.

The oldest fossil molluscs date from 570 MYA (*e.g.* Runnegar & Pojeta, 1974, 1985; Valentine, 1980), near the Precambrian-Cambrian boundary. This period was marked by an explosive radiation of animals resulting in the appearance of most extant invertebrate phyla (*e.g.*

[1] Author for correspondence.

Table 1. List of 17 oligonucleotides complementary to conserved regions in eukaryotic 18S rRNA genes. These were used to determine the sequence of both strands of the 18S rRNA gene of *Onchidella celtica*.

Sequence[1]	Strand[2]	Corresponding position in the 18S rRNA gene of *Onchildella celtica*
CTGGTTGATYCTGCCAGT	R	4–21
GAAACTGCGAATGGCTCATT	R	82–101
AATGAGCCATTCGCAGTTTC	C	101–82
AGGGYTCGAYYCCGGAGA	R	393–410
TCTCCGGRRTCGARCCCT	C	410–393
TCTCAGGCTCCYTCTCCGG	C	422–404
ATTACCGCGGCTGCTGGC	C	605–588
CGCGGTAATTCCAGCTCCA	R	597–615
TTGGYRAATGCTTTCGC	C	990–974
TTRATCAAGAACGAAAGT	R	1002–1019
CCGTCAATTYYTTTRAGTTT	C	1188–1169
AATTTGACTCAACACGGG	R	1221–1238
GGGCATCACAGACCTGTTAT	C	1479–1460
ATAACAGGTCTGTGATGCCC	R	1460–1479
TTTGYACACACCGCCCGTCG	R	1666–1685
GACGGGCGGTGTGTRC	C	1684–1669
CYGCAGGTTCACCTACRG	C	1833–1816

[1] Sequence positions where both purines (A and G) are present are indicated by "R", those where both pyrimidines (C and T) are present by "Y".
[2] Oligonucleotides with a sequence corresponding to that of the RNA-like strand are indicated by a "R", those whose sequence is complementary to it, by a "C".

Bergström, 1991; Erwin, 1991; Valentine, 1991). Several aspects of the metazoan branching pattern still remain confused due to the doubtful homology of the relatively few morphological, anatomical and embryological characters shared by different phyla (*e.g.* Nielsen, 1977; Anderson, 1981; Inglis, 1985; Bergström, 1986; Ax, 1989; Schram, 1991; Backeljau *et al.*, 1993). Nevertheless, Mollusca appear to be a monophyletic group belonging to the Spiralia (*i.e.* Platyhelminthes, Nemertini, Mollusca, Sipuncula, Echiura and Annelida, and probably Gnathostomulida and Entoprocta) (*e.g.* Wingstrand, 1985; Brusca & Brusca, 1990; Willmer, 1990), but no synapomorphies are known linking the Mollusca unambiguously to any other spiralian phylum (*e.g.* Wingstrand, 1985; Erwin, 1991). Some authors (*e.g.* von Salvini-Plawen, 1990a) suggest a sister group relationship to Turbellaria (Platyhelminthes) considering the flat, often ciliated, ventral creeping foot as a synapomorphy relating both phyla. Many others however, include the Mollusca in the protostome clade (*e.g.* Wingstrand, 1985; Brusca & Brusca, 1990; Willmer, 1990; Schram, 1991).

Biochemical and molecular characters have been introduced as an independent source of phylogenetic information. A serological study of molluscs, echinoderms, annelids and arthropods suggested that Mollusca are most closely related to Annelida (Wilhelmi, 1944). Lyddiatt *et al.* (1978) used cytochrome *c* amino acid sequence data to deduce a sister group relationship between molluscs and echinoderms. In studies using 5S ribosomal RNA (rRNA) sequences (Ohama *et al.*, 1984; Hendriks *et al.*, 1986; Hori & Osawa, 1987), the Mollusca (represented by Bivalvia, Gastropoda and Cephalopoda) appeared as a polyphyletic group. From the analysis of Lenaers and Bhaud (1992) on the basis of partial sequences of 28S rRNA, Mollusca (represented by *Mytilus edulis*) appeared to be a sister group to Annelida. Holland *et al.* (1991) used partial small subunit (SSU) rRNA (18S rRNA) sequences and suggested that Mollusca (represented by *Mytilus edulis*) and Arthropoda are sister taxa. On the basis of mitochondrial SSU rRNA sequences the Mollusca, represented by a prosobranch and a chiton, appeared as sister group to the Annelida or as a paraphyletic group including the latter phylum (Ballard *et al.*, 1992) . In all these studies, however, the data sets were too limited to allow reliable conclusions. Field *et al.* (1988) determined partial sequences of SSU rRNA from representatives of ten different metazoan phyla including four Mollusca, viz. an opisthobranch gastropod, two bivalves and a chiton. Yet, different phylogeny inference methods yielded contradictory results. Field *et al.* (1988; see also Raff *et al.*, 1989) used a distance method to conclude that Mollusca form a clade with Annelida, Sipuncula, Brachiopoda and Pogonophora. However, the relationships between the five groups were not resolved. Ghiselin (1988, 1989) reanalyzed this data set with a "signature" approach and concluded that molluscs are a sister group to the Annelida *sensu lato* (*i.e.* Annelida *sensu strictu*, Brachiopoda, Pogonophora and Sipuncula). A maximum parsimony analysis of the same data produced a similar clade containing Sipuncula, Pogonophora, Brachiopoda, Annelida and Mollusca (Patterson, 1989) but with the latter two phyla not being monophyletic. Lake (1989), who applied evolutionary parsimony, also concluded that Mollusca are paraphyletic.

In a preliminary attempt, we use complete SSU rRNA sequences to assess molluscan phylogeny. We consider sequences to be complete if (1) the sequence of the entire 18S rRNA molecule is known or (2) if only a total number on the order of 50 nucleotides at the 5′ and 3′ terminal parts are missing because they are used as PCR primer annealing sites (*e.g.* Rice, 1990; Littlewood, 1991). Hitherto, complete 18S rRNA sequences of only three molluscan species viz., the bivalves *Placopecten magellanicus* and *Crassostrea virginica* and the gastropod *Limicolaria kambeul*, have been published respectively by Rice (1990), Littlewood (1991) and Winnepenninckx *et al.* (1992) . In this paper we present the complete 18S rRNA sequence of the gymnomorphan snail *Onchidella celtica* (Cuvier, 1817). A fifth molluscan sequence (Winnepenninckx *et al.* , 1993), that of the chiton *Acanthopleura japonica* (Lischke, 1873) is also included.

Small ribosomal subunit RNA sequences

SSU rRNA sequences combine several features that make them appropriate for phylogenetic studies (Raff *et al.*,

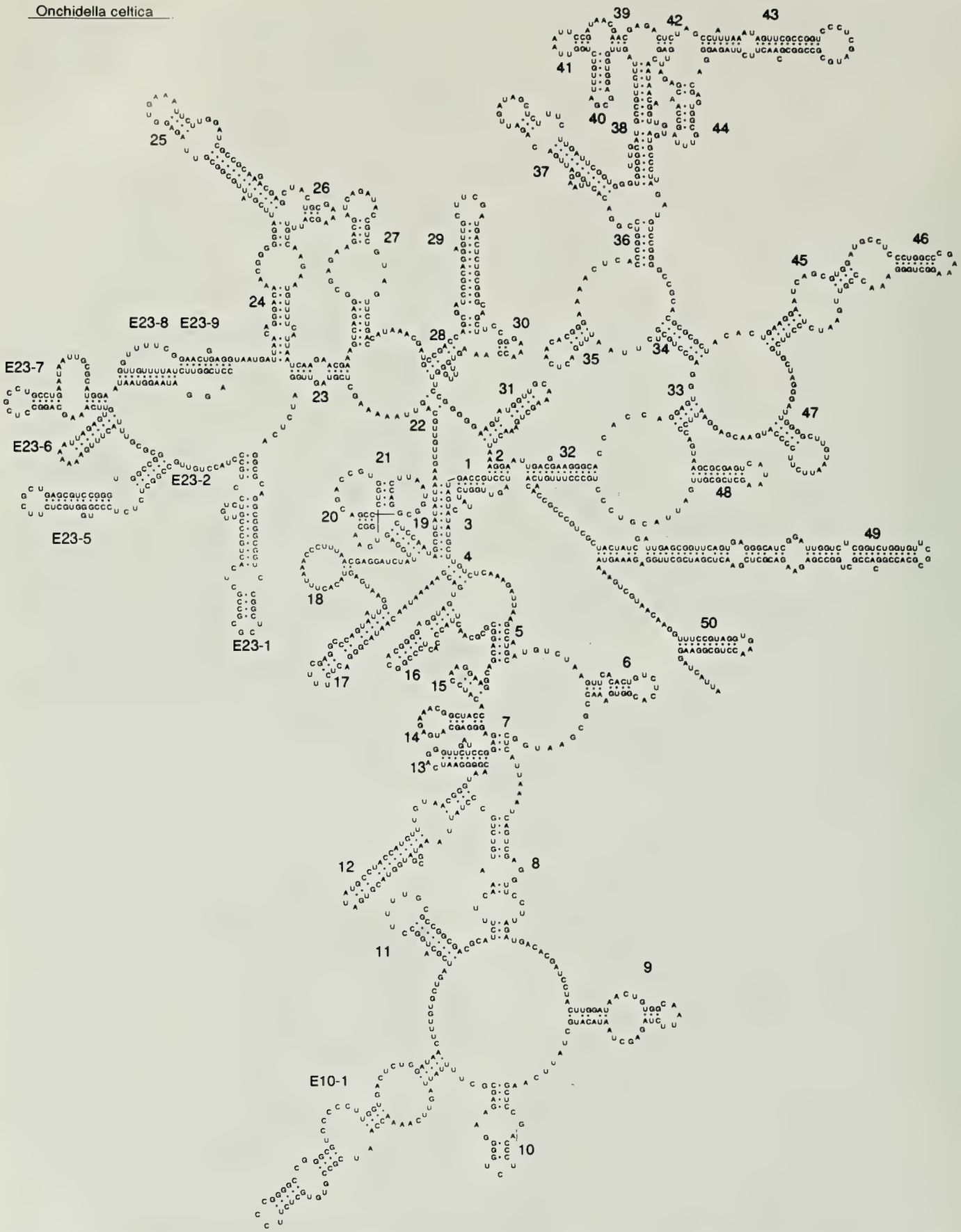


Figure 1. Secondary structure model for the 18S rRNA of *Onchidella celtica*. Helix numbering from helix 19 onward has changed with respect to the numbering used by De Rijk *et al.* (1992), due to the discovery of a tertiary structure interaction in helix 19 (Woese & Gutell, 1989).

Table 2. The 18S rRNA sequence of the gastropod *Onchidella celtica*.

0001	UAUCUGGUUGAUCCUGCCAGUAGUCAUAUGCUUGUCUCAAAGAUUAAGCC
0051	AUGCAUGUCUAAGUUCACACUGUCUCACGGUGAAACCGCGAAUGGCUCAU
0101	UAAAUCAGUCGAGGUUCCUUAGAUGACACGAUCCUACUUGGAUAACUGUG
0151	GCAAUUCUAGAGCUAAUACAUGCUAUUCAAGCUCCGACCCUCUGGGGAAG
0201	AGCGCUUUUAUUAGUUCAAAACCAAUCGCCGUGUGCUCUUCCCGGGGCCG
0251	GGCGUCCCCCUUGGUGACUCUGGAUAACUUUGUGCUGAUCGCAUGGCCUU
0301	UUGCGCCGGCGACGCAUCUUUCAAAUGUCUGCCCUAUUAAAUGCGAUGGU
0351	ACGUGAUAUGCCUACCAUGUUUGUAACGGGUAACGGGGAAUCAGGGUUCG
0401	AUUCCGGAGAGGGAGCAUGAGAAACGGCUACCACAUCCAAGGAAGGCAGC
0451	AGGCGCGCAACUUACCCACUCCCGGCACGGGGAGGUAGUGACGAAAAAUA
0501	ACAAUACGGGACUCUUUCGAGGCCCAGUAAUUGGAAUGAGUACACUUUAA
0551	ACCCUUUAACGAGGAUCUAUUGGAGGGCAAGUCUGGUGCCAGCAGCCGCG
0601	GUAAUUCCAGCUCCAAUAGCGUAUAUUAAAGUUGUUGCAGUUAAAAAGCU
0651	CGUAGUUGGAUCUCAGGCGCAGGCGGGCGGUCCGGCUCGCGCCGCUCACU
0701	GCCCGUUGUCUCCUGCCCUACCUGUUGCCGGCUCUCUCCCGUGGGUGCUC
0751	UUCGCUGAGCGUCCGGGUGGCCGGCGCGUUUACUUUGAAAAAAUUAGAGU
0801	GUUCAAAGCAGGCCUCGCCUGCCUGAAUAAUUGCGCAUGGAAUAAUGGAA
0851	UAGGACCUCGGUUCUAUUUUGUUGGUUUUCGGAACUGGAGGUAAUGAUUA
0901	ACAGGGACAAACGGGGGGAUUCGUAUUGCGGCGUUAGAGGUGAAAUUCUU
0951	GGAUCGCCGCAAGACGAGCUACUGCGAAAGCAUUUGUCAAGAAUGUUUUC
1001	AUUAAUCAAGAACGAAAGUCAGAGGCGAGAAGACGAUCAGAUACCGUCGU
1051	AGUUCUGACCAUAAACGAUGCCGACCAGCGAUCCGCAGGAGUUGCUUCGA
1101	UGACUCUGCGGGCAGCUUCCGGGAAACCAAAGUGUUUGGGUUCCGGGGGA
1151	AGUAUGGUUGCAAAGCUGAAACUUAAAGGAAUUGACGGAAGGGCACCACC
1201	AGGAGUGGAGCCUGCUGCUUAAUUUGACUCAACACGGGAAAACUCACCCG
1251	GUCCGGACACUGUAAGGAUUGACAGAUUGAUAGCUCUUUCUUGAUUCGGU
1301	GGGUGGUGGUGCAUGGCCGUUCUUAGUUGGUGGAGCGAUUUGUCUGGUUA
1351	AUUCCGAUAACGAACGAGACUCUAGCCUAUUAAAUAGUUCGCCGGUCCCU
1401	CGAUGCGCCGGCGCAACUUCUUAGAGGGACGAGUGGCGUUUAGCCAACGA
1451	GAUUGAGCAAUAACAGGUCUGUGAUGCCCUUAGAUGUCCGGGGCCGCACG
1501	CGCGCUACACUGAAGGAAUCAGCGUGGAUGCCUCCCUGGCCCGAAAGGCU
1551	GGGAAACCCGUUGAAUCUCCUUCGUGCUAGGGAUUGGGGCUUGUAAUUCU
1601	UCCCCAUGAACGAGGAAUUCCCAGUAAGCGCGAGUCAUAAGCUCGCGUUG
1651	AUUACGUCCCUGCCCUUUGUACACACCGCCCGUCGCUACUAUCGAUUGAG
1701	CGGUUCAGUGAGGGCAUCGGAUUGGUCUCGGUCUGGUGUUCGCGCACCGG
1751	CACCGCUGGCCGAGAAGACGCUCGAACUCGAUCGCUUGGAGAAAGUAAAA
1801	GUCGUAACAAGGUUUCCGUAGGUGAACCUGCGGAAGGAUCAUUA

1989; Hillis & Dixon, 1991; Solignac *et al.*, 1991; Woese, 1991): (1) universality; (2) constancy of function; (3) alternation of conserved regions with variable ones, allowing phylogenetic studies at a broad range of taxonomical levels; (4) presence of conservative regions that allow the design of "universal" primers; (5) a conservative secondary structure facilitating the identification of homologous positions in regions with little sequence similarity; (6) apparent absence of lateral gene transfer; (7) a large information content (1800–1900 bp) (8) intraspecific sequence homogeneity among different gene copies (Gerbi, 1985; Dover, 1986).

Gene cloning

Much sequence information on rRNAs has been obtained by direct RNA sequencing using reverse transcriptase (Lane *et al.*, 1985; Solignac *et al.*, 1991) or by direct sequencing of the rRNA genes after PCR amplification (Saiki *et al.*, 1988). Both techniques are very rapid. Yet we prefer to clone and sequence the 18S rRNA genes, for direct RNA sequencing has some disadvantages: (1) RNA is less stable than DNA; (2) subsequent checking of sequences is not possible; (3) sequencing of regions with strong secondary structure is difficult; (4) reverse transcriptase has a rather high error frequency; (5) only one strand is available and thus two-strand verification is not possible. All this results in an overall error rate of about 1% (Lane *et al.*, 1985). Although PCR amplification eliminates a great deal of these problems, it also has some drawbacks (Hillis & Dixon, 1991): (1) *Taq* polymerase has a high error rate, viz. $\approx 2\times10^{-4}$ to $<1\times10^{-5}$ according to Eckert and Kunkel (1991) and $2.75\ 10^{-3}$ according to Bej *et al.* (1991); (2) the 3′ and 5′ parts of the gene itself have to be used as primer annealing sites, if the sequence of the adjacent regions is unknown; (3) direct sequencing of PCR amplified fragments is difficult (*e.g.* Gyllenstein, 1989); (4) the product is afterwards not available to others for verification. By cloning the PCR product prior to sequencing, the latter two problems can be overcome, but the sequencing of numerous clones is necessary to avoid an enhancement of the error rate

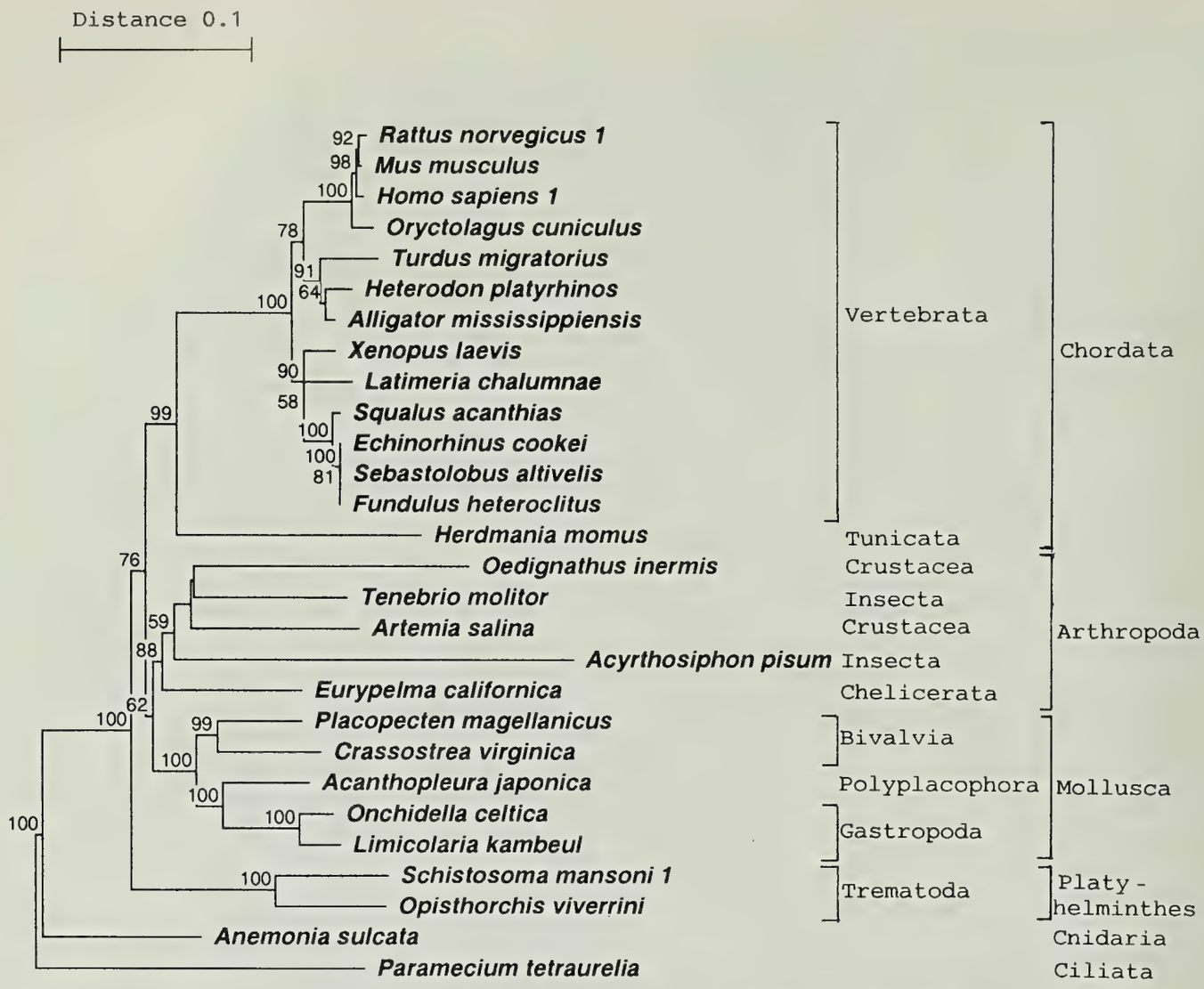


Figure 2. Neighbor-joining tree based on the 18S rRNA sequences from 27 Metazoa. All sequences were complete except for the following (number of sequenced nucleotides between brackets): *Turdus migratorius* (1753), *Alligator mississippiensis* (1691), *Heterodon platyrhinos* (1717) and *Latimeria chalumnae* (1777). *Paramecium tetraurelia* was chosen as an outgroup. Bootstrap values are indicated at the root of each clade, but only if they exceed 50%.

(Bevan *et al.*, 1992). This of course reduces the time advantage of PCR amplification.

MATERIALS AND METHODS

Animals: Specimens of *Onchidella celtica* collected at Vila Franca do Campo (São Miguel, Azores) were frozen alive and preserved at –80°C. Voucher material was deposited in the collections of the "Koninklijk Belgisch Instituut voor Natuurwetenschappen", Brussels (general inventory number, I.G. No. 28053).

DNA extraction: Digestive glands of ten specimens were pooled and homogenized under liquid nitrogen in a prechilled mortar and transferred to 15 ml of preheated (60°C) 2% CTAB buffer (2% (w/v) CTAB; 0.2% (v/v) 2-mercaptoethanol; 1.4 M NaCl; 20 mM EDTA; 100 mM Tris-HCl pH=8; 100 μg/ml proteinase K). After incubation at 60°C for 30 min., further extraction was done

Table 3. Organisms that were used as outgroup in our analyses.

Species	Position
Zea mays	angiosperms
Neurospora crassa	ascomycetes
Saccharomyces cerevisiae	ascomycetes
Rhodosporidium toruloides	basidiomycetes
Gracilaria lemaneiformis	red algae
Porphyra umbilicalis	red algae
Chlorella ellipsoidea	green algae
Volvox carteri	green algae
Prorocentrum micans	dinoflagellates
Giardia duodenalis	diplomonads
Trypanosoma brucei	kinetoplastids
Paramecium tetraurelia	ciliates
Oxytricha nova	ciliates
Plasmodium berghei	apicomplexa (Sporozoa)
Dictyostelium discoideum	slime molds

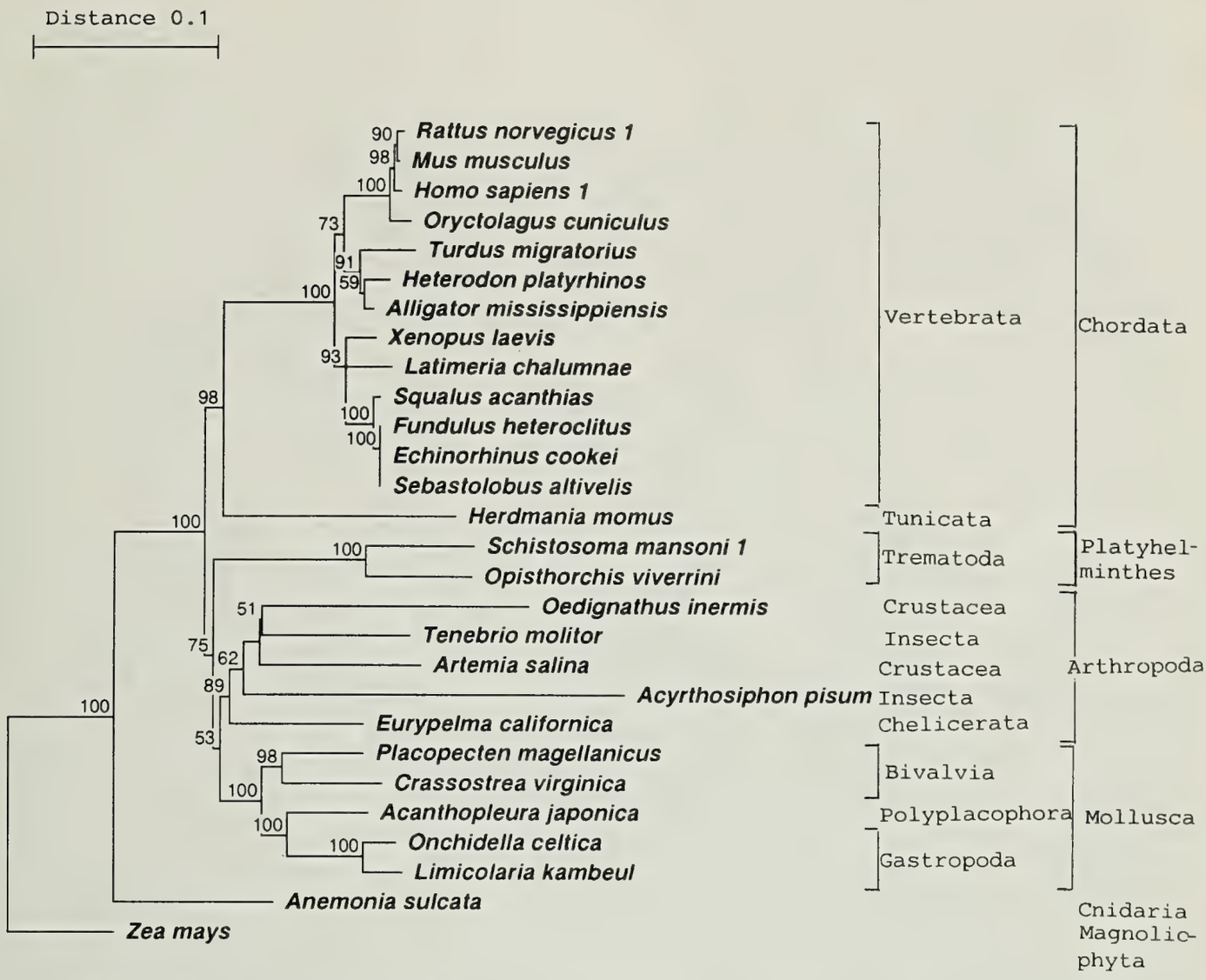


Figure 3. Neighbor-joining tree based on the same set of metazoan 18S rRNA sequences as in Fig. 2, but with *Zea mays* as an outgroup. Bootstrap values are indicated as in Fig. 2.

as described by Winnepenninckx *et al.* (1993a). The DNA yield amounted to 60 μg.

Gene cloning and sequencing: Restriction enzymes suitable for isolation of a DNA fragment containing the 18S rRNA gene were identified as described by Winnepenninckx *et al.* (1992). After digestion of 1.2 μg DNA with *Bam*HI and separation on a 0.8% (w/v) agarose gel, restriction fragments of 4 kb containing the 18S rRNA gene were eluted (Heery *et al.*, 1990). Competent DH5α *E. coli* cells (Gibco BRL Life Technologies; Gaithersburg, USA) were transformed with these DNA restriction fragments ligated into pBluescriptSK$^+$ (Stratagene; La Jolla, California, USA). Colony screening was performed using a PCR fragment of the gastropod *Limicolaria kambeul* (Winnepenninckx *et al.*, 1992), labeled with ^{32}P via nick translation (Rigby *et al.*, 1977). Plasmids were isolated (Birnboim & Doly, 1979) from a single clone and sequencing was performed by the dideoxynucleotide method (Sanger *et al.*, 1977) using Sequenase 2.0 (USB; Cleveland, Ohio, USA). The 18S rRNA primers used are given in Table 1.

Sequence alignment and construction of phylogenetic trees: The *Onchidella celtica* 18S rRNA sequence was aligned with other SSU rRNA sequences present in our database (De Rijk *et al.*, 1992). Alignment was done manually taking into account the secondary structure features of the molecule, as described by De Rijk *et al.* (1992). For tree construction, pairwise distances were calculated using the formula of Jukes and Cantor (1969) modified to take into account gaps (Van de Peer *et al.*, 1990). They served to derive neighbor-joining trees (Saitou & Nei, 1987), whose reliability was tested by bootstrapping (Felsenstein, 1985) over 100 replicates. According to the guidelines of Hillis and Bull (1993), only branching points with bootstrap values higher than 70% were considered to be reliable. Estimated internal branches with bootstrap values above 70% should represent true clades over 95% of the time (Hillis & Bull,

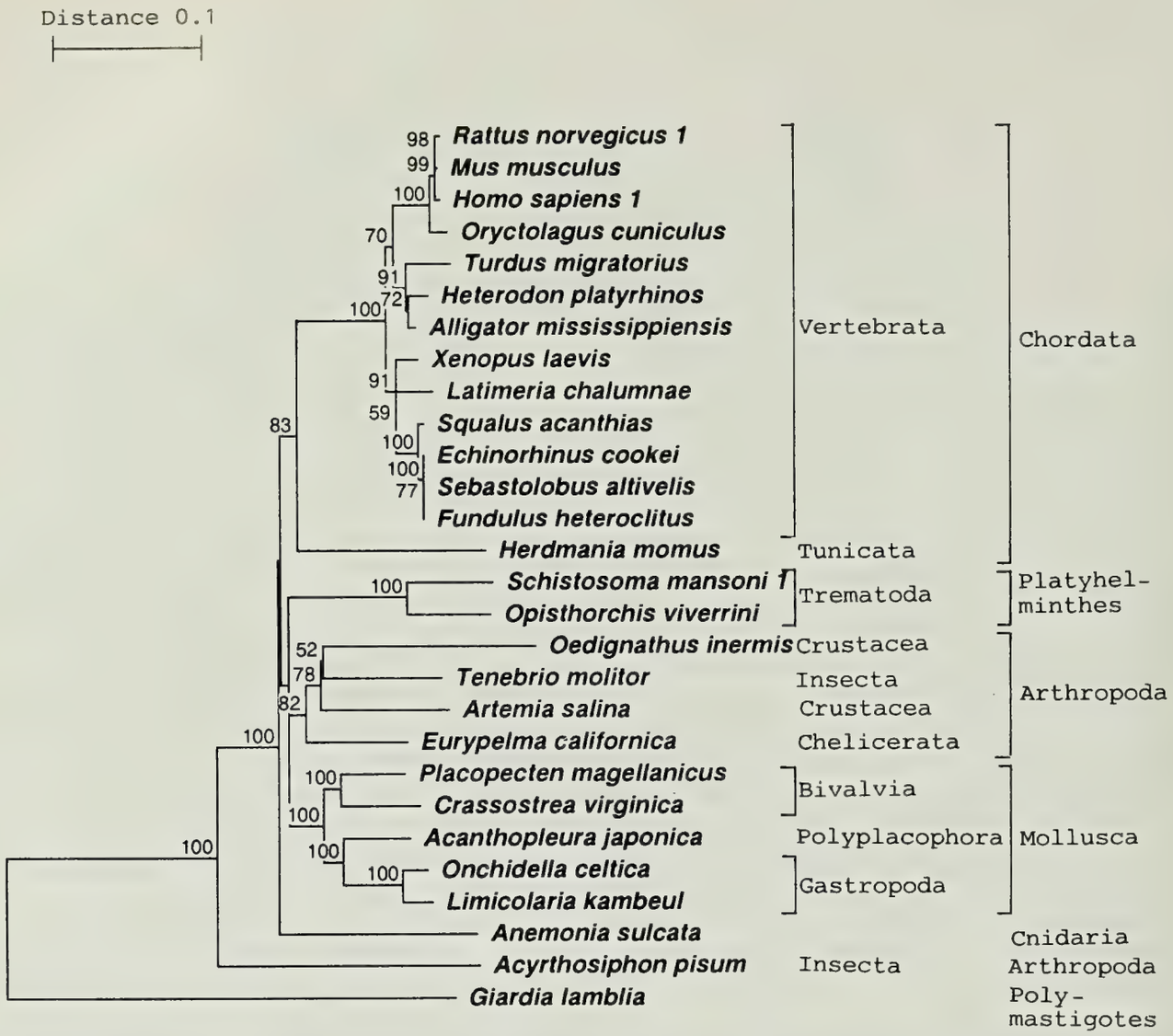


Figure 4. Neighbor-joining tree based on the same set of metazoan 18S rRNA sequences as in Fig. 2, but with *Giardia duodenalis* (often called *Giardia lamblia* or *Giardia intestinalis*) as an outgroup. Bootstrap values are indicated as in Fig. 2.

1993). All calculations were carried out with the TREECON package of Van de Peer and De Wachter (1993). Character state analyses using maximum parsimony were performed using the package HENNIG86 (version 1.5; Farris, 1989) with the heuristic algorithms MHENNIG* and BB* combined. The results were summarized in a strict consensus tree, *i.e.* a tree that contains only those clusters that are common to all competing trees ("nelsen" command of HENNIG86). Nucleotides were treated as non-additive characters and no differential weighting was done.

RESULTS

Sequence Alignment

The 18S rRNA of *Onchidella celtica* (EMBL accession number X70211), of which the nucleotide sequence is shown in Table 2, is 1844 nucleotides long. The 3′ and 5′ termini of the gene were located on the basis of similarity with those of other 18S rRNA sequences. Figure 1 shows a secondary structure model of the molecule in accordance with the one published for *Limicolaria kambeul* (Winnepenninckx *et al.*, 1992). Both models show high similarity to each other and are in accordance with the general model proposed for eukaryotic SSU rRNA (De Rijk *et al.* 1992). Based on our latest insights into the secondary structure of 18S RNA, modifications were made in helices 19, 20, 21 and 38. The new gastropod sequence as well as the one of *Acanthopleura japonica* (Winnepenninckx *et al.*, 1993b) were added to an alignment of other SSU rRNA sequences (De Rijk *et al.*, 1992). This alignment can be obtained on request. Trees were constructed on the basis of a set of 27 metazoan sequences which are either complete or nearly complete.

Distance Matrix Analyses

Figure 2 shows the neighbor joining (NJ) tree obtained with the ciliate *Paramecium tetraurelia* as outgroup. It

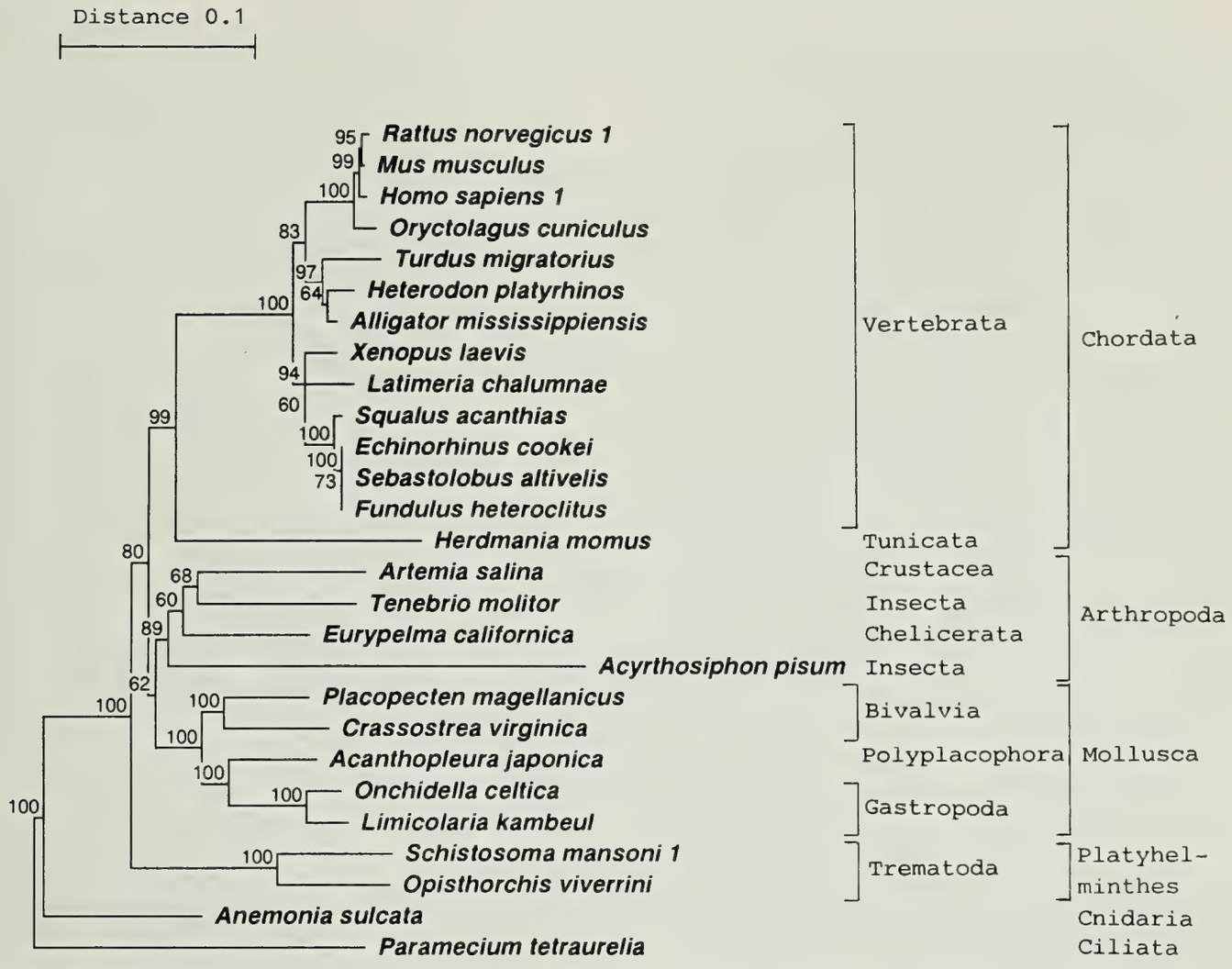


Figure 5. Neighbor-joining tree obtained on the basis of a set containing all the 18S rRNA metazoan sequences of Fig. 2 except *Oedignatus inermis*. *Paramecium tetraurelia* was used as an outgroup. Bootstrap values are indicated as in Fig. 2.

suggests that (bootstrap values in parentheses): (1) Mollusca are a monophyletic group (100/100) within a relatively poorly supported protostome clade (62/100); (2) Gastropoda (100/100) and Bivalvia (99/100) are monophyletic as well; (3) Polyplacophora appears as a sister group to the Gastropoda (100/100). The tree also indicates that : (1) Cnidaria are a sister group to Eubilateria (100/100); (2) Acoelomata, represented by two Trematoda, are a sister group to the Eucoelomata (76/100); (3) Arthropoda are a monophyletic group (88/100); (4) neither Insecta nor Crustacea are monophyletic; (5) Chordata (99/100) and Vertebrata (100/100) are both monophyletic.

We attempted to assess the stability of our tree by testing its sensitivity to the presence of specific taxa. First we studied the influence of the outgroup by successively replacing *Paramecium tetraurelia* by each of the 14 other organisms listed in Table 3. We observed only two topological changes. In nine out of the 15 cases, the topology shown in Figure 3 was obtained, *i.e.* the Platyhelminthes appeared as a sister group to the Arthropoda-Mollusca clade. In one case, the topology shown in Figure 4 was obtained, viz. when the diplomonad *Giardia duodenalis*, was chosen as outgroup. This organism forms a very long branch in previously published trees comprising organisms from different eukaryotic kingdoms (*e.g.* Van de Peer *et al.*, 1993). In this case, the aphid *Acyrthosiphon pisum*, which is also marked by an exceptionally long branch, became a sister group to all other Metazoa. The latter observation is probably due to the fact that errors in distance estimation increase with the amount of divergence. Long branches will provoke an underestimation of the evolutionary distance and will systematically attract each other, causing biased topologies (Felsenstein, 1978; Olsen, 1987; Swofford & Olsen, 1990; Lake, 1991). The changes in the position of the Platyhelminthes, which do not have exceptionally long branches, is probably not due to such a systematic error. The low bootstrapping values on their branching point, suggests uncertainty as to their position. Inclusion of representatives of more invertebrate phyla might be helpful in this case.

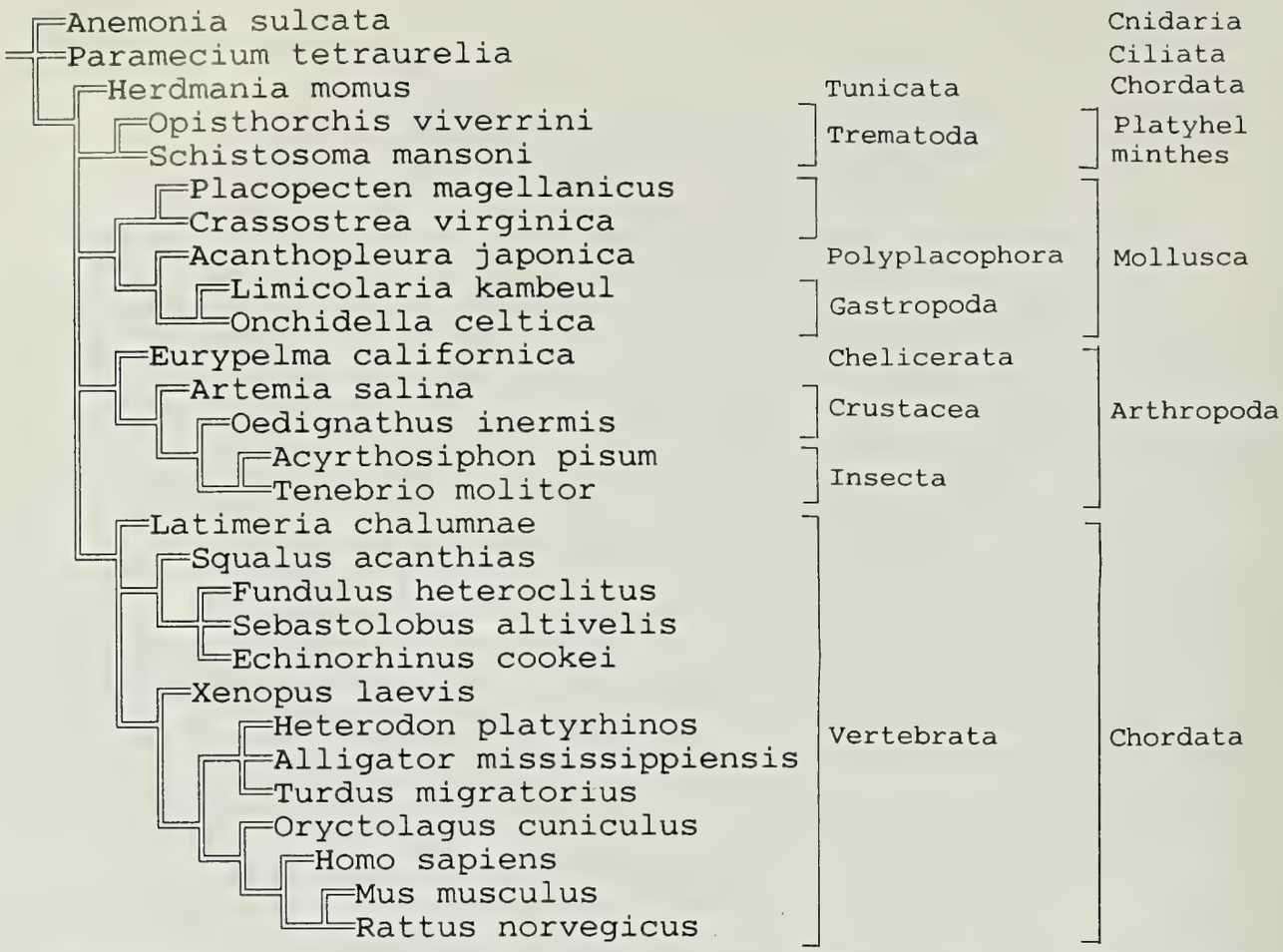


Figure 6. Strict consensus tree constructed from three maximum parsimony trees (length=3114 steps; c.i.= 0.51) obtained by applying the MHENNIG*+BB* option of Hennig86 on the 706 informative positions of the same alignment as in Fig. 2 and with *Paramecium tetraurelia* chosen as outgroup.

Subsequently, we constructed 27 trees with *Paramecium tetraurelia* as outgroup, but each time omitting one species. Only one topological change was observed: when excluding *Oedignathus inermis*, *Acyrtosiphon pisum* branched off first within the arthropod clade (Figure 5). The fact that this change involves the species with the longest branch, again points to the above mentioned "long branch effect" (Felsenstein, 1978; Swofford & Olsen, 1990). Since the placement of the two Platyhelminthes was ambiguous (*cfr.* Figures 2 and 3) and since we suspected *Acyrthosiphon pisum* to be a source of systematic errors, we removed all three species from our data set to assess their impact. However, the topology of the tree we obtained did not differ from the one in Figure 2.

Character State Analyses

The 28 species analysed, with *Paramecium tetraurelia* as outgroup, yielded 706 informative sites. A position is informative if it contains at least two different nucleotides, each of them present in at least two species (Nei, 1987). Ambiguous nucleotides were not used to ascertain the informative character of a position. Three maximum parsimony (MP) trees of 3114 steps and with a consistency index (c.i.) (Kluge & Farris, 1969) of 0.51 were found. The strict consensus tree shown in Figure 6 suggests that (1) Mollusca, Bivalvia and Gastropoda are monophyletic groups; (2) Polyplacophora appear as a sister group to Gastropoda; (3) Arthropoda are a monophyletic clade in which Chelicerata branch off first; (4) Insecta are monophyletic but Crustacea are paraphyletic; (5) Vertebrata are monophyletic. Ten different data input orders did not change this topology. Again we tested the stability of our results. If the placement of a taxon is biased, its removal should cause an increase of the consistency index (Swofford & Olsen, 1990). We checked this by successively removing those species, the position of which appeared unstable in our distance matrix analyses, viz. *Acyrthosiphon pisum*, *Schistosoma mansoni*, *Opisthorchis viverrini*, and a fourth species, *Mus musculus*, which occupied a stable position. Each time we identified the informative positions anew and applied HENNIG86 with *Paramecium tetraurelia* as outgroup. Omitting *Acyrthosiphon pisum* increased the c.i. to 0.53, while removing any of the other species did not change the c.i. This again suggests that the placement of *Acyrthosiphon pisum* is liable to a systematic error. As for the ambiguous position of the Platyhelminthes, this may

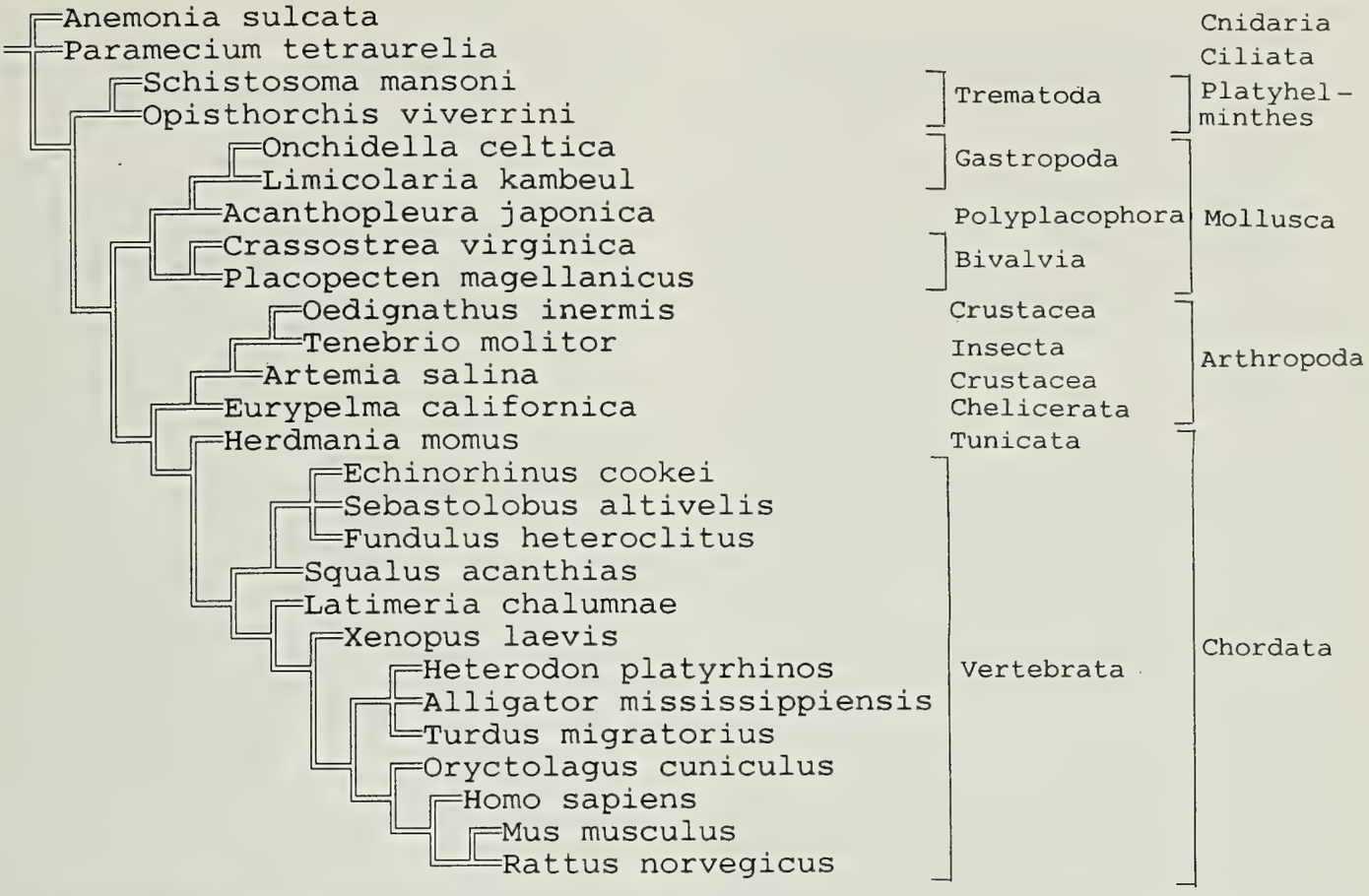


Figure 7. Strict consensus tree constructed from two maximum parsimony trees (length=2741 steps; c.i.=0.53) obtained with the MHENNIG*+BB* option on the 658 informative sites of the same alignment of Fig. 2 from which the insect *Acyrthosiphon pisum* was removed.

be due to the lack of other invertebrate phyla and classes. Figure 7 shows the strict consensus tree of the two MP trees (length=2741; c.i.=0.53) obtained when *Acyrthosiphon pisum* was excluded. All our conclusions based on the tree in Figure 6 remain valid, but in addition the bilaterian pentachotomy of Figure 6 is now resolved. Mollusca appear as a sister group to a clade containing Arthropoda and Chordata. It is also suggested that (1) Bilateria are monophyletic; (2) Acoelomata are a sister group to Eucoelomata; (3) Chordata are monophyletic.

DISCUSSION

The monophyletic character of the Mollusca, the Bivalvia and the Gastropoda, which is supported by all our trees, is generally accepted (*e.g.* Brusca & Brusca, 1990; Willmer, 1990; von Salvini-Plawen, 1985, 1990a; Götting, 1980). Using globin amino acid sequences, Goodman *et al.* (1988) agreed with these views. The 5S rRNA based analyses of Ohama *et al.* (1984), Hendriks *et al.* (1986) and Hori and Osawa (1987) also confirmed gastropod monophyly. Ghiselin (1988, 1989) supported molluscan monophyly. But Patterson (1989) and Lake (1989) did not corroborate these well established views, while the question was not resolved by Field *et al.* (1988; see also Raff *et al.*, 1989).

In both the distance and MP trees, we find the chiton included within the conchiferan clade as a sister group to the Gastropoda. This result is in contrast with the results of anatomical (*e.g.* Milburn, 1960; Stasek, 1972; Götting, 1980; Scheltema, 1988; Brusca & Brusca, 1990; von Salvini-Plawen, 1990a) and paleontological (*e.g.* Runnegar & Pojeta, 1974; Pojeta, 1980; Peel, 1991) studies. Neither Field *et al.* (1988; see also Raff *et al.*, 1989), nor Ghiselin (1988, 1989) or Lake (1989) were able to resolve the position of the Polyplacophora, while Patterson (1989) suggested that Polyplacophora and Brachiopoda are sister taxa. Using mitochondrial SSU rRNA sequences (Ballard *et al.*, 1992), the class either appeared as a sister group to the Gastropoda-Annelida clade or formed together with the Gastropoda a sister group to the Annelida. Addition of more molluscan representatives to our data set is necessary to investigate the conflicting position of the Polyplacophora.

Our current data set is also not sufficiently representative to draw conclusions on the position of the Mollusca among other Metazoa. From our NJ analyses, the phylum appears as a sister group to the Arthropoda (see also

Holland *et al.*, 1991) but this topology is insufficiently supported by bootstrap values. According to the character state analysis it branches off before the Chordata-Arthropoda clade. Data from additional invertebrate phyla should be included.

All our current analyses strongly support the view that: (1) Arthropoda is a monophyletic group and (2) Vertebrata, Chordata and Bilateria are monophyletic. These observations are in agreement with the results of some classical (*e.g.* Ax, 1989; Brusca & Brusca, 1990; Schram, 1991) and molecular studies (Ghiselin, 1988, 1989; Patterson, 1989; Raff *et al.*, 1989; Winnepenninckx *et al.*, 1992). Contradictory views on these aspects of metazoan phylogeny were given by *e.g.* Lake (1989), Willmer (1990) and Fryer (1992). However our analyses do not allow conclusions on the status of the Acoelomata and the mono- or paraphyletic character of the Insecta and Crustacea. It is beyond the scope of this paper to expand on metazoan evolution, however the congruence of most of our results with independently derived hypotheses suggests that complete 18S rRNA sequences are a reliable tool to assess the phylogeny of the Mollusca and other metazoan groups.

ACKNOWLEDGMENTS

We are indebted to Dr. A.M. Frias Martins (University of the Azores) for helping to collect *Onchidella celtica*. The investigations were performed in the framework of the Institute for the Study of Biological Evolution of the University of Antwerp. This work was supported by FKFO Grants 2.0023.94 and 2.0003.93. B. Winnepenninckx holds an I.W.O.N.L. scholarship.

LITERATURE CITED

Anderson, D. T. 1981. Origins and relationships among animal phyla. Proceedings of the Linnean Society of New South Wales 106:151–166.

Ax, P. 1989. Basic phylogenetic systematization of the Metazoa. *in*: Fernholm, B., K. Bremer and H. Jörnvall (eds). The Hierarchy of life. Elsevier, Amsterdam, p. 229–245.

Backeljau, T., B. Winnepenninckx and L. De Bruyn. 1993. Cladistic analysis of metazoan relationships: areappraisal. Cladistics 9:167–181.

Ballard, J. W. O., G. J. Olsen, D. P. Faith, W. A. Odgers, D. M. Rowell and P. W. Atkinson. 1992. Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods. Science 258:1345–1348.

Bej, A. K., M. H. Mahbubani and R. M. Atlas. 1991. Amplification of nucleic acids by polymerase chain reaction (PCR) and other methods and their applications. Critical reviews in Biochemistry and Molecular Biology 26:301–334.

Bergström, J. 1986. Metazoan evolution - a new model. Zoologica Scripta 15:189–200.

Bergström, J. 1991. Metazoan evolution around the Precambrian-Cambrian transition. *In*: Simonetta, A.M. and S.C. Morris (eds). The early evolution of Metazoa and the significance of problematic taxa. Cambridge University Press, p. 25–34.

Bevan, I. S., R. Rapley and M. R. Walker. 1992. Sequencing of PCR-amplified DNA. PCR Methods and Applications 1:222–228.

Birnboim, H. C. and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Research 7:1513–1522.

Brusca, R. C. and G. J. Brusca. 1990. Invertebrates. Sinauer Associates, Sunderland, MA 922 pp.

De Rijk, P., J.-M. Neefs, Van de Peer, Y. and R. De Wachter. 1992. Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Research 20:2075–2089.

Dover, G. A. 1986. Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends in Genetics 2:159–165.

Eckert, K. A. and T. A. Kunkel. 1991. DNA polymerase fidelity and the polymerase chain reaction. PCR Methods and Applications 1:17–24.

Erwin, D. H. 1991. Metazoan phylogeny and the Cambrian radiation. Trends in Ecology and Evolution 6: 131–134.

Farris, J. S. 1989. Hennig86: a PC-DOS program for phylogenetic analysis. Cladistics 5:163.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27:401–410.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Field, K. G., G. J. Olsen, D. J. Lane, S. J. Giovannoni, M. T. Ghiselin, E. C. Raff, N. R. Pace and R. A. Raff. 1988. Molecular phylogeny of the animal kingdom. Science 239: 748–753.

Fryer, G. 1992. The origin of the Crustacea. Acta Zoologica 73:273–286.

Gerbi, S. A. 1985. Evolution of ribosomal DNA. *In*: MacIntyre, R.J. (ed). Molecular Evolutionary Genetics. Plenum Press, New York and London, p. 419–517.

Ghiselin, M. T. 1988. The origin of molluscs in the light of molecular evidence. *In*: Harvey, P. H. and L. Partridge (eds). Oxford Surveys in Evolutionary Biology. Volume 5. p. 66–95.

Ghiselin, M. T. 1989. Summary of our present knowledge of metazoan phylogeny. *In*: Fernholm, B., K. Bremer and H. Jörnvall (eds). The Hierarchy of life. Elsevier, Amsterdam, p. 261–272.

Goodman, M., J. Pedwaydon, J. Czelusniak, T. Suzuki, T. Gotoh, L. Moens, F. Shishikura, D. Walz and S. Vinogradov. 1988. An evolutionary tree for invertebrate globin sequences. Journal of Molecular Evolution 27:236–249.

Götting, K.-J. 1980. Origin and relationships of the Mollusca. Zeitschrift für zoologische Systematik und Evolutionsforschung 18:24–27.

Gyllenstein, U. B. 1989. PCR and DNA sequencing. Biotechniques 7:700–708.

Heery, D. M., F. Gannon and R. Powell. 1990. A simple method for subcloning DNA fragments from gel slices. Trends in Genetics 6:173.

Hendriks, L., E. Huysmans, A. Vandenberghe and R. De Wachter. 1986. Primary structures of the 5S ribosomal RNAs of 11 arthropods and applicability of 5S RNA to the study of metazoan evolution. Journal of Molecular Evolution. 24:103–109.

Hillis, D. M. and M. T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. The Quarterly Review of Biology. 66:411–453.

Hillis, D. M. and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology. 42:182–192.

Holland, P. W. H., A. M. Hacker and N. A. Williams. 1991. A molecular analysis of the phylogenetic affinities of *Saccoglossus cambrensis* Brambell & Cole (Hemichordata). Philosophical Transactions of the Royal Society of London (Series B) 332:185–189.

Hori, H. and S. Osawa. 1987. Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. Molecular Biology and Evolution 4:445–472.

Inglis, W. G. 1985. Evolutionary waves: patterns in the origins of animal phyla. Australian Journal of Zoology 33:153–178.

Jukes T. H. and C. R. Cantor. 1969. Evolution of protein molecules. *In*: Munro, H. N. (ed). Mammalian Protein Metabolism. Academic Press, New York, p. 21–132.

Kluge, A. G. and J. S. Farris. 1969. Quantitative phyletics and the evolution of anurans. Systematic Zoology 18:1–32.

Lake, J. A. 1989. Origin of the multicellular animals. *In*: Fernholm, B., K. Bremer and H. Jörnvall (eds). The Hierarchy of life. Elsevier, Amsterdam, p. 273–278.

Lake, J. A. 1991. Tracing origins with molecular sequences: metazoan and eukaryotic beginnings. Trends in Biochemical Sciences 16:46–50.

Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin and N. R. Pace. 1985. Rapid determination of 16S-ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Science of the USA 82:6955–6959.

Lenaers, G. and M. Bhaud. 1992. Molecular phylogeny of some polychaete annelids: an initial approach to the Atlantic-Mediterranean speciation problem. Journal of Molecular Evolution 35:429–435.

Littlewood, D. T. J., S. E. Ford and D. Fong. 1991. Small subunit rRNA gene sequence of *Crassostrea virginica* (Gmelin) and a comparison with similar sequences from other bivalve molluscs. Nucleic Acids Research 19:6048.

Lyddiatt, A., D. Peacock and D. Boulter. 1978. Evolutionary change in invertebrate cytochrome C. Journal of Molecular Evolution 11:35–45.

Milburn, P. W. 1960. Further remarks on the interpretation of the Mollusca. The Veliger 3:43–48.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, p. 315–319.

Nielsen, C. 1977. The relationships of Entoprocta, Ectoprocta and Phoronida. American Zoologist 17:149–150.

Ohama, T., T. Kumazaki, H. Hori and S. Osawa. 1984. Evolution of multicellular animals as deduced from 5S rRNA sequences: a possible early emergence of the Mesozoa. Nucleic Acids Research 12:5101–5108.

Olsen, G. J. 1987. Earliest phylogenetic branchings: comparing rRNA-based evolutionary trees inferred with various techniques. Cold Spring Harbor Symposia on Quantitative Biology LII:825–837.

Patterson, C. 1989. Phylogenetic relations of major groups: conclusions and prospects. *In*: Fernholm, B., K. Bremer, and H. Jörnvall (eds). The Hierarchy of life. Elsevier, Amsterdam, p. 471–487.

Peel, J. S. 1991. Functional morphology of the class Helcionelloida nov., and the early evolution of the Mollusca. *In*: Simonetta, A.M. and S.C. Morris (eds). The early evolution of Metazoa and the significance of problematic taxa. Cambridge University Press, p. 157–177.

Pojeta, J. Jr. 1980. Molluscan phylogeny. Tulane Studies in Geology and Paleontology 16:55–80.

Raff, R. A., K. G. Field, G. J. Olsen, S. J. Giovannoni, D. J. Lane, M. T. Ghiselin, N. R. Pace and E. C. Raff. 1989. Metazoan phylogeny based on analysis of 18S ribosomal RNA. *In*: Fernholm, B., K. Bremer and H. Jörnvall (eds). The Hierarchy of life. Elsevier, Amsterdam, p. 247–260.

Rice, E. L. 1990. Nucleotide sequence of the 18S ribosomal RNA gene from the Atlantic sea scallop *Placopecten magellanicus*. Nucleic Acids Research 18:5551.

Rigby, P. W. J., M. Dieckmann, C. Rhodes and P. Berg. 1977. Labeling deoxyribonucleic acids to high specific activity in vitro by nick translation with DNA polymerase I. Journal of Molecular Biology 113: 237–251. Runnegar, B. and J. Pojeta, Jr. 1974. Molluscan phylogeny: the paleontological viewpoint. Science 186:311–317.

Runnegar, B. and J. Pojeta, Jr. 1985. Origin and diversification of the Mollusca. *In*: Wilbur, K. M., E. R.

Trueman and M. R. Clarke (eds). The Mollusca. Volume 10. Evolution. Academic Press, Inc. p. 1–57.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. Erlich. 1988. Primer-directed enzymatic amplification of

DNA with a thermostable DNA polymerase. Science 239:487–491.

Saitou, N. and M. Nei. 1987. The Neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406–425.

Sanger, F., S. Nicklen and R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Science of the USA 74:5463–5467.

Scheltema, A. H. 1988. Ancestors and descendents: relationships of the Aplacophora and Polyplacophora. American Malacological Bulletin 6:57–68.

Schram, F. R. 1991. Cladistic analysis of metazoan phyla and the placement of fossil problematica. *In*: Simonetta, A. M. and S. C. Morris (eds). The early evolution of Metazoa and the significance of problematic taxa. Cambridge University Press, p. 35–46.

Solignac, M., M. Pélandakis, F. Rousset and A. Chenuil. 1991. Ribosomal RNA phylogenies. *In*: Hewitt G.M., A. W. B. Johnston and J. P. W. Young (eds). NATO ASI Series. Volume 57. Molecular Techniques in Taxonomy. Volume H57. Springer-Verlag, Berlin, Heidelberg, p. 73–85.

Stasek, C. R. 1972. The molluscan framework. *In*: Florkin, M. and B. T. Scheer (eds). Chemical Zoology. Volume VII. Mollusca. Academic Press, New York and London, p. 1–44.

Steiner, G. 1992. Phylogeny and classification of Scaphopoda. Journal of Molluscan Studies 58:385–400.

Swofford, D. L. and G. J. Olsen. 1990. Phylogeny reconstruction. *In*: Hillis, D. M. and C. Moritz (eds). Molecular Systematics. Sinauer Associates, Inc., Sunderland, MA, p. 411–501.

Valentine, J. W. 1980. L'origine des grands groupes d'animaux. La Recherche 112:666–674.

Valentine, J. W. 1991. Major factors in the rapidity and extant of the metazoan radiation during the Proterozoic-Phanerozoic transition. *In*: Simonetta, A.M. and S.C. Morris (eds). The early evolution of Metazoa and the significance of problematic taxa. Cambridge University Press, p. 11–13.

Van de Peer, Y., J.-M. Neefs and R. De Wachter. 1990. Small ribosomal subunit RNA sequences, evolutionary relationships among different life forms, and mitochondrial origins. Journal of Molecular Evolution 30:463–476.

Van de Peer, Y., J.-M. Neefs, P. De Rijk and R. De Wachter. 1993. Evolution of eukaryotes as deduced from small

ribosomal subunit RNA sequences. Biochemical Systematics and Ecology 21:43–56.

Van de Peer, Y. and R. De Wachter. 1993. TREECON: A software package for the construction and drawing of evolutionary trees. Computer Applications in the Biosciences 9: 177–182.

von Ihering, H. 1876. Versuch eines natürlichen Systemes der Mollusken. Jahrbuch der Deutschen Malakozoologischen Gesellschaft 3: 97–148.

von Salvini-Plawen, L. 1969. Solenogastres und Caudofoveata (Mollusca, Aculifera) Organisation und phylogenetische Bedeutung. Malacologia 9:191–216.

von Salvini-Plawen, L. 1972. Zur Morphologie und Phylogenie der Mollusken; die Beziehungen der Caudofoveata und der Solenogastres als Aculifera, als Mollusca und als Spiralia. Zeitschrift zur Wissenschaftlichen Zoologie 184: 205–394.

von Salvini-Plawen, L. 1985. Early evolution and the primitive groups. *In*: Wilbur, K. M., E. R. Trueman and M. R. Clarke (eds). The Mollusca. Volume 10. Evolution. Academic Press, Inc. p. 59–150.

von Salvini-Plawen, L. 1990a. Origin, phylogeny and classification of the phylum Mollusca. Iberus 9:1–33.

von Salvini-Plawen, L. 1990b. The status of the Caudofoveata and the Solenogastres in the mediterranean sea. Lavori della Società Italiana di Malacologia 23: 5–30.

Wilhelmi, R. W. 1944. Serological relationships between the Mollusca and other invertebrates. Biological Bulletin 87: 96–105.

Willmer, P. 1990. Invertebrate relationships—Patterns in animal evolution. Cambridge University Press, Cambridge.

Wingstrand, K. G. 1985. On the anatomy and relationships of recent Monoplacophora. Galathea Reports 16:7–94.

Winnepenninckx, B., T. Backeljau, Y. Van de Peer and R. De Wachter. 1992. Structure of the small ribosomal subunit RNA of the pulmonate snail *Limicolaria kambeul*, and phylogenetic analysis of the Metazoa. FEBS Letters 309: 123–126.

Winnepenninckx, B., T. Backeljau and R. De Wachter. 1993a. Extraction of high molecular weight DNA from molluscs. Trends in Genetics 9:407.

Winnepenninckx, B., T. Backeljau and R. De Wachter. 1993b. Complete small ribosomal subunit RNA sequence of the chiton *Acanthopleura japonica* (Lischke, 1873) (Mollusca, Polyplacophora). Nucleic Acids Research 21:1670.

Woese, C. R. 1991. The use of ribosomal RNA in reconstructing evolutionary relationships among Bacteria. *In*: Selander, R. K., A. G. Clark and T. S. Whittam (eds). Evolution at the molecular level. Sinauer Associates, Inc., Sunderland, MA p. 1–25.

Woese, C. R. and R. R. Gutell. 1989. Evidence for several higher order structural elements in ribosomal RNA. Proceedings of the National Academy of Science of the USA 86:3119–3122.

Preliminary Ribosomal RNA Phylogeny of Gastropod and Unionoidean Bivalve Mollusks

Gary Rosenberg
Academy of Natural Sciences
1900 Benjamin Franklin Parkway
Philadelphia, PA 19103 USA

Gerald S. Kuncio
Department of Medicine
University of Pennsylvania
Philadelphia, PA 19104 USA

George M. Davis
Academy of Natural Sciences
1900 Benjamin Franklin Parkway
Philadelphia, PA 19103 USA

M. G. Harasewych
Department of Invertebrates
National Museum of Natural History
Smithsonian Institution
Washington, D.C. 20560 USA

ABSTRACT

Sequences of about 150 nucleotides in the D6 region of the large (28S) ribosomal molecule were obtained from 20 unionoidean bivalves and 13 gastropods, including 9 truncatellids, 1 muricid, 1 cancellariid, 1 melongenid, and 1 pleurotomariid. These were analyzed along with sequences from Emberton *et al.* (1990) for 8 pulmonates, a helicinid and a pomatiopsid. Rates of divergence varied by a factor of two, with unionoidean and the helicinid sequences differing by 15–20% relative to mouse, pulmonates by 22–24%, neogastropods by 24–27%, and rissooideans by 27–32%. Length variation among sequences occurred mainly in the D6 loop, and complementary mutations were seen in the D6 stem. Cladistic analysis found 24 equally parsimonious trees; the strict consensus supports monophyly of Unionoidea, Rissooidea, Pulmonata and Stylommatophora. Monophyly of Neogastropoda is not contradicted.

Some groupings are anomalous when compared to morphology-based phylogenies. Helicinidae groups with Unionoidea, Pleurotomariidae with Neogastropoda, and *Geomelania* (Truncatellidae) with Pomatiopsidae. In each case, addition of taxa that intersect long branches (e.g. Chitonidae, Patellogastropoda) might show that characters interpreted as synapomorphic are plesiomorphic or convergent. The observed grouping of Muricidae and Cancellariidae is well-supported, indicating that cancellariids are a highly derived group within the Stenoglossa. Pleurotomariidae are more closely related to the other gastropods in the analysis than are Helicinidae, supporting Haszprunar's (1988) anatomy-based conclusion.

To date, sequence studies of mollusks have not overturned phylogenies based on morphology, but rather have helped in choosing among competing morphology-based hypotheses. Like morphological data, sequence data are subject to problems of convergence, unequal rates of evolution, and choice of taxa. Classifications must be based on all available data to maximize the potential for detecting convergences and correctly resolving phylogenetic relationships.

Key words: Gastropoda, Bivalvia, 28S ribosomal RNA, phylogeny, cladistics, rates of divergence.

INTRODUCTION

Ribosomal RNA (rRNA) and rDNA sequences have proven to be valuable and versatile sources of data for phylogenetic inferences. The great variation in rates of evolution of different parts of rDNA allows evolutionary investigations from the level of population and species through kingdom, by study of appropriately variable regions. This variation has caused debate as to the reliability of some types of rDNA sequence data in phylogenetic analysis, but it has become clear that, when analyzed with care, all parts of the sequence are potentially informative (Wheeler & Honeycutt, 1988; Swofford & Olsen, 1990; Hillis & Dixon, 1991; Dixon & Hillis, 1993). Ribosomal sequences have most often been used in determining relationships among bacteria (e.g., Woese & Olsen, 1986) and vertebrates (e.g., Hedges *et al.*, 1990), but can be used with any organism (e.g., Sogin *et al.*, 1986; Field *et al.*, 1988).

Of the more than 150 phylogenetic studies of rDNA sequences published to date (Hillis & Dixon, 1991), only a few have been devoted to mollusks. Ghiselin (1988) looked at molluscan origins using 18S rRNA, Emberton *et al.* (1990) at pulmonate relationships using D6 28S rRNA and Tillier *et al.* (1992) at gastropod phylogeny using D1 28S rRNA. These studies have demonstrated the potential for rDNA sequences to sharpen and resolve ideas of molluscan phylogeny, particularly at the ordinal level and above.

We have supplemented the 10 gastropod sequences obtained by Emberton *et al.* (1990) with data from 33 more molluscan species. All 43 sequences were used in this study with the aims of 1) analyzing aspects of caenogastropod, archaeogastropod, and unionoidean relationships, 2) surveying variability in 28S rRNA sequences in mollusks, and 3) examining how choice of taxa affects

Table 1. Localities and catalogue numbers of voucher specimens for this study. Depository is the Academy of Natural Sciences of Philadelphia (ANSP) unless otherwise noted; USNM = United States National Museum. Condition: f = frozen; l = lyophilized; w = whole live animal. Voucher lot of *Anodonta imbecillis* was collected in 1974; tissue sample was from same population in 1975. See Emberton *et al.* (1990) for vouchers of *Helicina orbiculata, Oncomelania hupensis, Biomphalaria glabrata, Mesodon inflectus, Mesodon normalis, Neohelix albolabris, Triodopsis hopetonensis, Haplotrema concavum, Mesomphix latior, Ventridens cerinoideus.*

Species	Condition	Locality	Catalogue no.
Anodonta cataracta	l	NE of Swedesboro, Gloucester Co., New Jersey	333526, 341937
	l	Swartswood, Sussex Co., New Jersey	334429, 341946
A. grandis	l	Ramah Borrow Canal, Iberville Parish, Louisiana	341888
A. imbecillis	l	Magnolia Springs, Jenkins Co., Georgia	333563
Amblema plicata	f	Bogue Chitto Creek, Dallas Co., Alabama	373820, A12742
Elliptio complanata	l	Swartswood, Sussex Co., New Jersey	334428
	l	Lake Lacawac, Lake Ariel, Wayne Co., Pennsylvania	339430
	l	Deep Creek, Nanticoke River, Sussex Co., Delaware	339340
Fusconaia cerina	f	Bogue Chitto Creek, Dallas Co., Alabama	397248
Gonidea angulata	l	Pit River, SW of Canby, Modoc Co., California	339965
Lampsilis teres	f	Bogue Chitto Creek, Dallas Co., Alabama	373821, A12744
L. claibornensis	f	Bogue Chitto Creek, Dallas Co., Alabama	397249
Megalonaias boykiniana	l	Ochlockonee River, Leon Co., Florida	346111
Obliquaria reflexa	f	Bogue Chitto Creek, Dallas Co., Alabama	397247, A12722
Quadrula cylindrica	l	Kyles Ford, Clinch River, Hancock Co., Tennessee	335041
Q. quadrula	f	Bogue Chitto Creek, Dallas Co., Alabama	397246, A12728
Plectomerus dombeyanus	l	Ramah Borrow Canal, Iberville Parish, Louisiana	vouchers not kept
Pleurobema cordatum	l	Ouachita River, Arkadelphia, Clark Co., Arkansas	340629
Unio pictorum	f	Shropshire Canal, north of Chester, near Mollington Grange, England	350622
Uniomerus tetralasmus	l	Magnolia Springs, Jenkins Co., Georgia	353133
Cumberlandia monodonta	l	Kyles Ford, Clinch River, Hancock Co., Tennessee	341956
Margaritifera falcata	l	Siletz River, Lincoln Co., Oregon	339339
M. margaritifera	l	Locust Creek, Schuylkill River, Pennsylvania	334867
Perotrochus maureri	f	90 miles east of Charleston, South Carolina	USNM 875218
Truncatella sp.	w	Harrison Point Lighthouse, Barbados	397286
T. caribaeensis	w	Bay side, Mile 57, Grassy Key, Florida Keys	397275
T. clathrus	w	Bay side, Mile 57, Grassy Key, Florida Keys	397273
T. pulchella	w	Bay side, Mile 57, Grassy Key, Florida Keys	397274
	w	Falmouth, Trelawny Parish, Jamaica	397264
T. reclusa	w	Cumaca, Northern Range, Trinidad	397285
T. scalaris	w	Falmouth, Trelawny Parish, Jamaica	397263
T. subcylindrica	w	The Fleet, Dorset, England	397280
Geomelania sp.	w	North of Quickstep, Trelawny Parish, Jamaica	397283
G. typica	w	Wallingford, St. Elizabeth Parish, Jamaica	397284
Busycon carica	f	Cape Henlopen, Sussex Co., Delaware	USNM 847010
Mancinella deltoidea	f	South Beach, Miami, Dade Co., Florida	USNM 870850
Progabbia cooperi	f	Off La Jolla, San Diego Co., California	USNM 846054

phylogenetic inference. The results presented here must be considered preliminary until data for longer sequences and additional ordinal level taxa are available.

MATERIALS AND METHODS

We obtained sequences from 20 unionoidean bivalves and 13 gastropods, including 9 truncatellids, 1 muricid, 1 cancellariid, 1 melongenid, and 1 pleurotomariid. Species names, localities, voucher information, and higher classification are given in Tables 1 and 2. Truncatellid RNA was obtained by homogenizing live animals, other gastropod RNA from frozen tissue, and unionoidean RNA from lyophilized or frozen tissue. Methods for sequencing followed Emberton *et al.* (1990) and were done in the same laboratory, using the same primer for the D6 region, complementary to nucleotides 2099 through 2118 for mouse as published by Hassouna *et al.* (1984). Each species was sequenced at least twice, or more often as necessary to resolve ambiguities in nucleotide identity.

Sequence alignment: Gross alignment of the sequences was easily achieved because of large conservative stretches in the D6 flanks. The MALIGN program of Wheeler and Gladstein (version 1.73, 1993), was used to refine the manual alignment. The following weights (costs), with options alignaddswap and treeaddswap, yielded alignments that matched overall the manual alignments of the D6 flanks, while providing improvement in details:

transitions 1, transversions 3; internal gaps 10; leading gaps 5; trailing gaps 5.

Phylogenetic analysis: Informative and variable nucleotide positions, indicated by "i" or "v" in Figure 1, were analyzed using Hennig86. The sequence of commands "mhennig; bb; ie*;" was used, which guarantees finding all of the most parsimonious trees. All characters were equally weighted and unordered (command "cc-."). Those gaps marked with a hyphen (-) in Figure 1 were scored as characters, except for the deletion from positions 74 to 79 in *Truncatella clathrus*. Species that showed no differences in sequence were combined for the purpose of the phylogenetic analysis.

RESULTS

Aligned sequences are shown in Figure 1. The 5′ flanking region (positions 1–46), and the 3′ flank (positions 99–161) are conservative, and only a few gaps were inserted to align the sequences. The D6 loop shows considerable variation in length, and were too variable in the non-pulmonate gastropods to be reliably aligned. A number of complementary changes can be seen in the stem region, positions 47–55 and 90–98(Figure 2).

Among the 43 species, 26 different sequences were found. All species with identical sequences were confamilial. As reported by Emberton *et al.* (1990), sequences were invariant among the four polygyrids. Among 20 unionoideans, only 6 different sequences were found. *Cumberlandia*, *Gonidea* and the two *Margaritifera* all had distinct sequences. The other 16 unionids differed from each other by at most one nucleotide, falling into two groups, referred to here as the *Anodonta* and *Amblema* groups. In contrast, sequences differed strongly among truncatellids: each of the nine species had a unique sequence. Within the genus *Truncatella*, all species differed from each other by at least five nucleotides. Differences were concentrated in the D6 loop, and involved significant variation in length, in addition to nucleotide substitution. Only a partial sequence was obtained for *Perotrochus*.

Molluscan sequences differed from those of mouse at 15 to 32 percent of the sites (Table 3). Sequences from the unionoideans and *Helicina* were the most conservative, differing from mouse at 15 to 20 percent of sites. Gastropods, excluding *Helicina*, differed from mouse at 22 to 32 percent of sites, with pulmonates showing less sequence divergence (22 to 24 percent) than neogastropods (24 to 27 percent) and rissooideans (27 to 32 percent).

Use of published sequences (Gutell & Fox, 1988) from *Mus*, *Rattus*, *Xenopus* or *Homo* as outgroup did not affect polarization of characters. Sequences for *Caenorhabditis*, *Physarum* and *Saccharomyces* were more divergent from molluscan sequences than were vertebrate sequences, and in some regions could not be aligned satisfactorily with them. They were therefore judged less appropriate as outgroups. The sequence from *Drosophila* (Tautz *et al.*, 1988), shown in Figure 1, could be aligned, but differs from the molluscan sequences at twice as many sites as does the mouse sequence. Of 148 nucleotide positions scored in bivalves, 35 (24%) are variable relative to mouse, whereas 72 (49%) are variable relative to *Drosophila*. This degree of divergence made *Drosophila* unreliable as an outgroup.

Out of 153 alignable sites, 73 (48%) are variable in mollusks relative to mouse and 55 are potentially informative for cladistic analysis. With mouse as the outgroup, cladistic analysis yielded 24 equally parsimonious trees, the strict consensus tree of which is shown in Figure 3.

Table 2. Higher classification of genera for which sequence data were analyzed. Classification follows Davis and Fuller (1981) for Unionoidea, Haszprunar (1988b) for Gastropoda, Rosenberg (1989) for Rissooidea, Kantor and Harasewych (1992) for Neogastropoda and Emberton *et al.* (1990) for Pulmonata.

- Bivalvia
  - Paleoheterodonta
    - Unionoidea
      - Unionidae
        - Unioninae
          - *Unio*
        - Ambleminae
          - *Amblema*
          - *Megalonaias*
          - *Plectomerus*
          - *Quadrula*
        - Pleurobemini
          - *Elliptio*
          - *Fusconaia*
          - *Pleurobema*
          - *Uniomerus*
        - Gonideini
          - *Gonidea*
        - Lampsilini
          - *Lampsilis*
          - *Obliquaria*
        - Anodontinae
          - *Anodonta*
        - Margaritiferinae
          - *Margaritifera*
          - *Cumberlandia*
- Gastropoda
  - Neritopsina
    - Neritoidea
      - Helicinidae
        - *Helicina*
  - Vetigastropoda
    - Pleurotomarioidea
      - Pleurotomariidae
        - *Perotrochus*
  - Caenogastropoda
    - Neotaenioglossa
      - Rissooidea
        - Pomatiopsidae
          - *Oncomelania*
        - Truncatellidae
          - Truncatellinae
            - *Truncatella*
          - Geomelaniinae
            - *Geomelania*
    - Neogastropoda
      - Stenoglossa
        - Muricoidea
          - Muricidae
            - *Mancinella*
          - Melongenidae
            - *Busycon*
        - Cancellarioidea
          - *Progabbia*
- Pulmonata
  - Basommatophora
    - Planorboidea
      - Planorbidae
        - *Biomphalaria*
  - Stylommatophora
    - Holopoda
      - Polygyroidea
        - Polygyridae
          - *Mesodon*
          - *Neohelix*
          - *Triodopsis*
    - Holopodopes
      - Rhytidoidea
        - Haplotrematidae
          - *Haplotrema*
    - Aulacopoda
      - Zonitoidea
        - Zonitidae
          - *Mesomphix*
          - *Ventridens*

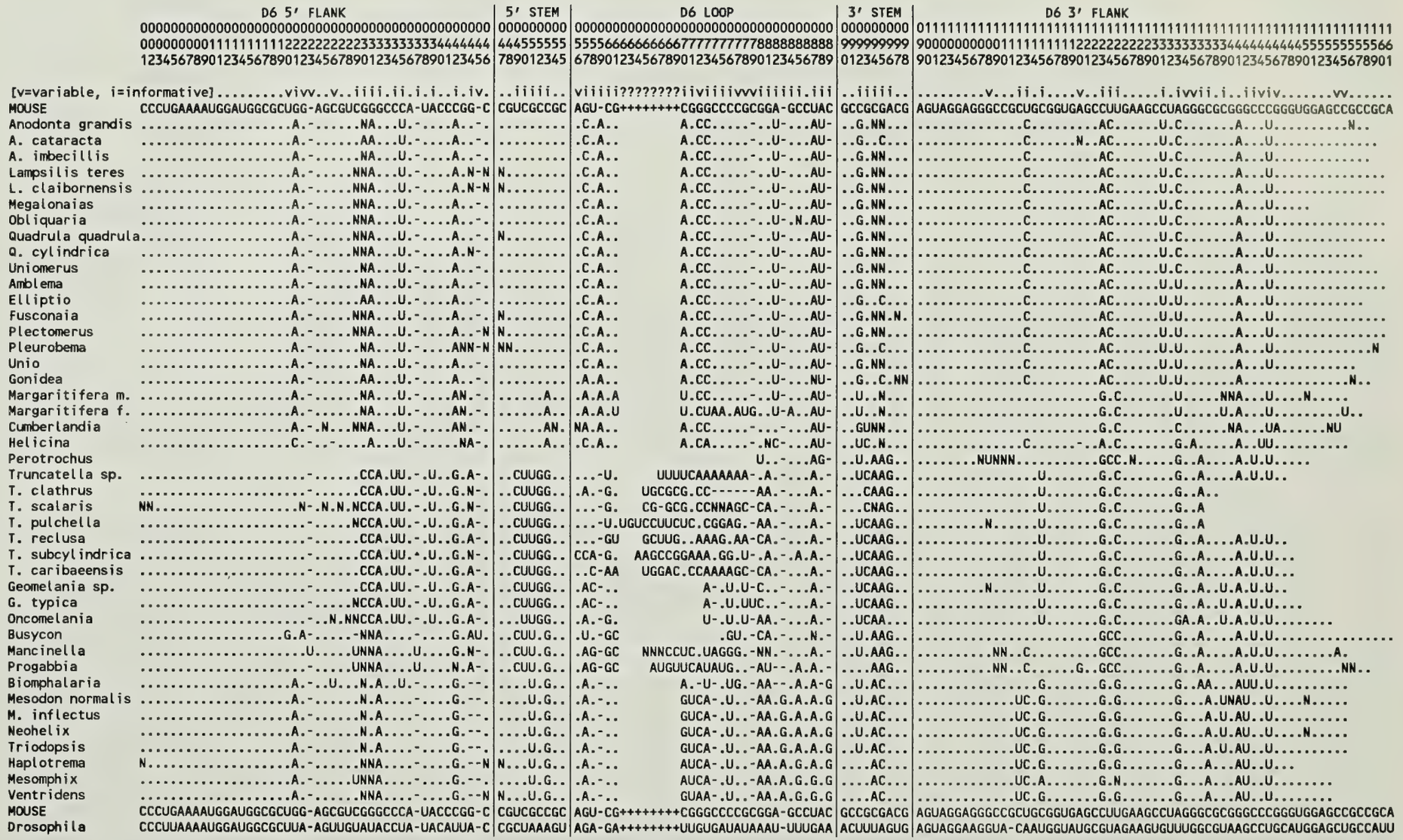


Figure 1. Aligned sequences. Numbers across the top in vertical triplets refer to the nucleotide position. The 5′ end of the sequences corresponds to mouse position 1836 of Hassouna *et al.* (1984). Nucleotides unchanged from mouse are represented by a dot (.); gaps by a dash (-). Nucleotides that could not be scored are marked "N". Pluses (+) in the D6 loop of mouse and *Drosophila* (at bottom) are an ellipsis where those sequences are longer than molluscan ones. Above the mouse sequence, positions variable in mollusks are marked "v"; those potentially informative for cladistic analysis are marked "i". The vertical lines delimit the 5′ stem, where the ribosomal molecule folds back and base pairs with itself (see Figure 2).

DISCUSSION

Analyses of rRNA data must take into account several complicating factors: the reliability of the alignments, bias in nucleotide composition, the ratio of transitions to transversions, and the affect of complimentary mutations in base-paired regions. The latter two are often handled by weighting the data in various ways. The significance of results, in terms of the reliability of nodes in a phylogram, is often assessed by bootstrapping.

Reliability of Alignments

Hillis and Dixon (1991) noted that alignments are often ambiguous when sequences differ by more than 30 percent in a given region. Except in the D6 loop, the molluscan sequences exhibited less than 30% divergence, and alignment was not problematic. In the phylogenetic analysis (Figure 3), all variable and informative nucleotide positions in the D6 loop were included, except numbers 62 to 69 (Figure 1). The alignment of these and of nucleotides 70 to 78 was uncertain in the rissooideans and neogastropods because of length polymorphism and sequence variation; most of these sequences are shown flush right with no gaps inserted. When characters 70 to 78 for these taxa were excluded from the analysis, the same consensus tree as shown in Figure 3 was obtained, except that all rissooideans formed a unresolved polychotomy, save the two *Geomelania*, which grouped together.

Bias in Nucleotide Composition

The sequences analyzed in this study exhibited a significant bias in favor of guanine and cytosine: the GC:AU ratio of mouse was 3:1, and all taxa had a ratio greater than 1.3:1 (Table 3). If mouse and mollusks had evolved this bias independently, it might affect the reliability of the mouse sequence as an outgroup for polarizing characters, because convergences could be mistaken for plesiomorphies. Because the sequences of mouse and mollusks have diverged only 15 to 30%, as discussed above, it is unlikely that the large GC biases evolved independently in these taxa. Nevertheless, as a control, the analysis was rerun with bivalves as outgroup, and with bivalves plus *Helicina* as outgroup. Topology of the ingroup was not affected by the change in outgroup.

Weighting of Character Data

Differential weighting schemes for character data are often used when analysis of a data set gives poorly resolved or unexpected phylogenetic inferences. If subsets of the data can be reasonably regarded as being less likely to suffer from homoplasy, they are given more weight. In rRNA sequences, transversions can be up to five times less frequent than transitions and are sometimes accorded greater weight (Crother & Presch, 1992). Mishler *et al.* (1988) suggested weighting according to the observed ratio of transitions to transversions in a given data set, but cautioned that this ratio is not likely uniform among taxa or in different regions within a sequence. Wheeler (1990) suggested a combinatorial weighting method based on observed nucleotide distributions among taxa, but this method assumes that there are no hidden intermediate states, and has been criticized on the grounds that it was applied *a posteriori* to trees based on equally weighted data (Albert & Mishler, 1992).

Taxon	000000000 444555555 789012345	000000000 999999999 012345678	paired sites
Mouse	CGUCGCCGC	GCCGCGACG	8
Anodonta group	CGUCGCCGC	GC**G**GCCACG	8
Amblema group	CGUCGCCGC	GC**G**GCCACG	8
Gonidea	CGUCGCCGC	GC**G**GCCANN	8
M. margaritifera	CGUCGC**A**GC	GCUGCNACG	8
M. falcata	CGUCGC**A**GC	GCUGCNACG	8
Cumberlandia	CGUCGCANC	GCGUNNACG	7?
Helicina	CGUCGC**A**GC	GCUCCNACG	7?
Perotrochus	?????????	GCUG**AAG**CG	?
Truncatella sp.	CG**CUUG**GGC	GCU**CAAG**CG	8.5
T. clathrus	CG**CUUG**GGC	GC**CCAAG**CG	9
T. scalaris	CG**CUUG**GGC	GC**CCNAG**CG	9
T. pulchella	CG**CUUG**GGC	GCU**CAAG**CG	8.5
T. reclusa	CG**CUUG**GGC	GCU**CAAG**CG	8.5
T. subcylindrica	CG**CUUG**GGC	GCU**CAAG**CG	8.5
T. caribaeensis	CG**CUUG**GGC	GCU**CAAG**CG	8.5
Geomelania sp.	CG**CUUG**GGC	GCU**CAAG**CG	8.5
G. typica	CG**CUUG**GGC	GCU**CAAG**CG	8.5
Oncomelania	CG**UUUG**GGC	GCU**CAAA**CG	8.5
Busycon	CG**CUU**CGGC	GCUG**AAG**CG	8.5
Mancinella	CG**CUU**CGGC	GCUG**AAG**CG	8.5
Progabbia	CG**CUU**CGGC	GC**CGAAG**CG	9
Biomphalaria	CGUC**U**CGGC	GCUG**A**CACG	7.5
Polygyridae	CGUC**U**CGGC	GCUG**A**CACG	7.5
Haplotrema	NGUC**U**CGGC	GC**CG**ACACG	8
Mesomphix	CGUC**U**CGGC	GC**CG**ACACG	8
Ventridens	NGUC**U**CGGC	GC**CG**ACACG	8

Figure 2. Detail of the D6 stem regions. Numbers across the top correspond to nucleotide positions in Figure 1. These regions fold back on each other (position 47 corresponding to 98, 48 to 97, etc.); base-pairs form where cytosine (C) matches guanine (G) and adenine (A) matches uracil (U). Mutations can disrupt the pairing, and complementary mutations sometimes restore the pairing. Inferred complementary mutations are in boldface. The number of paired sites is shown in the left column for each taxon. (The weaker pairing of uracil with guanine is counted as 0.5). Note that only a few species have all nine stem positions paired; complementary mutations do not always occur.

The observation of compensatory mutations in base-paired stem regions of rRNA molecules has led to the suggestion that such regions should be excluded from analyses, or weighted one-half because of non-independence of characters (Steele *et al.*, 1988; Wheeler & Honeycutt, 1988). In contrast, Smith (1989) found that results based on paired nucleotide positions were more reliable than those based on unpaired positions. Hillis and Dixon (1991) and Dixon and Hillis (1993) found that paired regions can contain most of the informative sites in some analyses, and suggested that paired regions be down-weighted by no more than twenty percent, because compensatory mutations do not always occur. This is true in the mollusk data, where perfect complementarity is not maintained (Figure 2).

We have opted to run our data with equal weights for

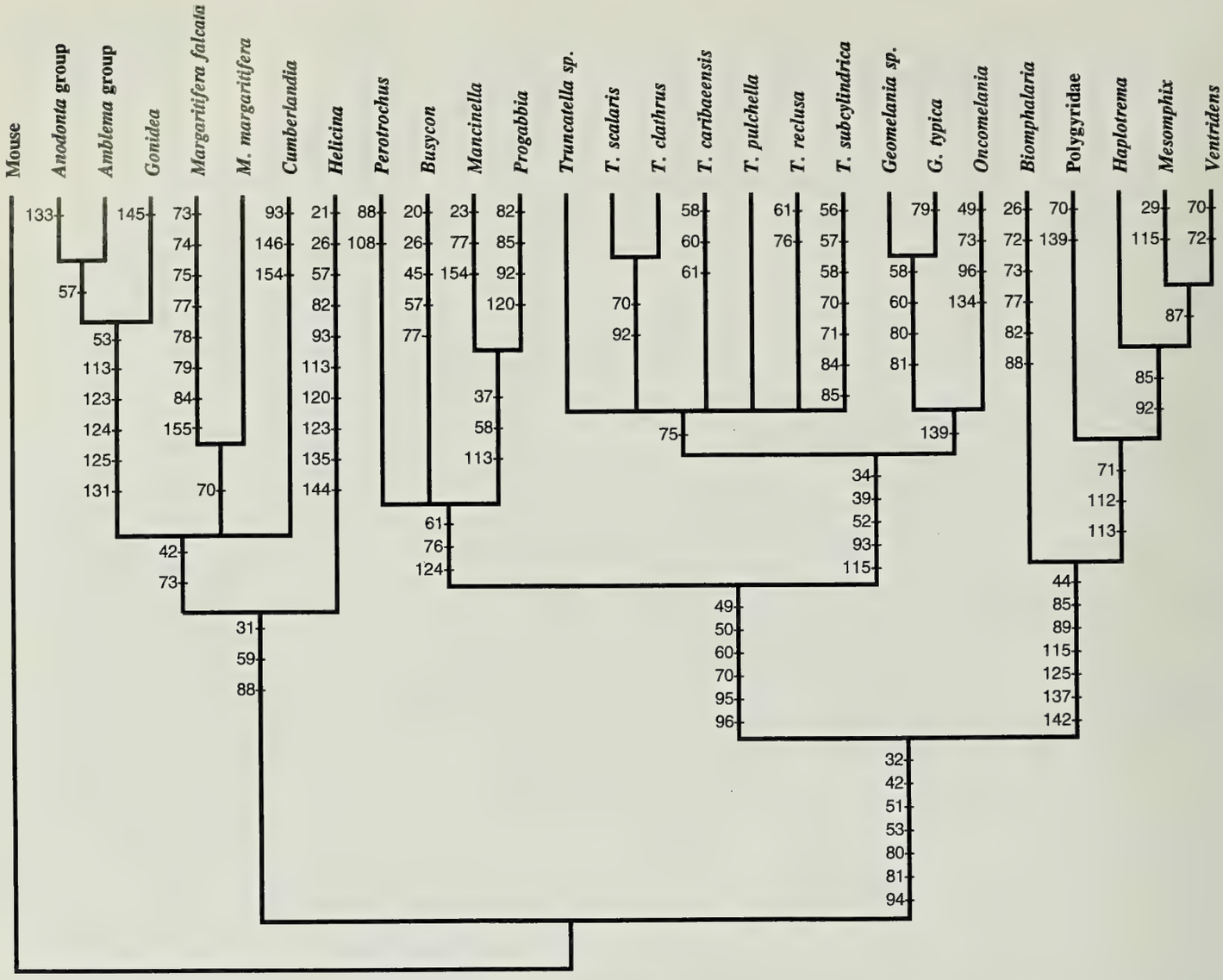


Figure 3. Strict consensus of 24 equally parsimonious trees of length 205 resulting from cladistic analysis of the sequence data. The consistency index of each of the most parsimonious trees is 0.64, the retention index is 0.81, and the rescaled consistency index is 0.52. Characters having states that define each node are indicated by nucleotide number. These characters support the nodes in each of the 24 most parsimonious trees making up the consensus tree, as shown using the character tracing feature of MacClade 3.01 (Maddison & Maddison 1992). Characters whose assignment to a node is ambiguous are not shown.

all characters for two reasons. First, because weighting schemes for transitions versus transversions and for compensatory mutations are still controversial, we prefer the assumptions of equal weighting to those that these schemes require. Second, the purpose of weighting is to compensate for the misleading effects of undetected homoplasy. (Homoplasy that has been detected is not misleading.) Undetected homoplasy exists when there are hidden intermediate character states between taxa that appear identical for a particular character. Given a sufficient number of undetected homoplasies, taxa that are convergent can be mistakenly inferred to be closely related. This is the long branch length problem first noted by Felsenstein (1978). One cannot know *a priori* what weights (if any) will overcome the long branch problem for a particular data set. Also, weighting cannot determine unknown intermediate states, so weighting cannot completely overcome the problem. Often, however, taxa can be added to the analysis that will intersect long branches, revealing the intermediate states. We therefore view convergence in sequence data primarily as a problem of taxon sampling rather than of weighting. In the following discussions, we attempt to identify areas where the addition of taxa is likely to change the relationships shown in Figure 3.

Reliability of Nodes

Hillis and Bull (1993) have recently shown that bootstrap proportions are conservative estimators of the probability of a clade's being real, if certain conditions are met. Hedges (1992) has shown that the number of bootstrap

Table 3. Differences in sequence and nucleotide composition between mollusks and mouse. If a position was not scored in a species, it was considered to match the consensus sequence for the higher taxon to which the species belongs. For example, at position 30 (figure 1), all unionoideans were considered to have "A," even though the nucleotide was not scorable in some species. Similarly, at the 3′ end, unionoidean sequences were all considered to extend to position 160, rissooideans to 153, neogastropods to 161, and pulmonates to 155. Positions 62 to 69 for mouse shown by pluses (+) in figure 1 are CGCGGCGU, extending from the 5′ side.

Sequence	Nucleotide differences	Total compared	Percentage difference	GC/AU
Mouse	—	—	—	3.1
Anodonta group	24	147	0.16	2.3
Amblema group	24	147	0.16	2.2
Gonidea	22	147	0.15	2.2
Margaritifera falcata	30	147	0.20	1.6
M. margaritifera	24	147	0.16	1.9
Cumberlandia	25	147	0.17	1.9
Helicina	27	139	0.19	2.1
Perotrochus	16	70	0.23	2.2
Truncatella sp.	42	143	0.29	1.4
T. clathrus	42	145	0.29	1.8
T. scalaris	41	145	0.28	1.9
T. pulchella	46	148	0.31	1.6
T. reclusa	43	145	0.30	1.6
T. subcylindrica	46	146	0.32	1.7
T. caribaeensis	46	145	0.32	1.7
Geomelania sp.	38	140	0.27	1.7
G. typica	39	140	0.28	1.6
Oncomelania	38	140	0.27	1.4
Busycon	36	149	0.24	1.9
Mancinella	40	152	0.26	1.9
Progabbia	41	153	0.27	1.8
Biomphalaria	34	142	0.24	1.6
Polygyridae	34	143	0.24	1.7
Haplotrema	33	143	0.23	1.8
Mesophix	32	143	0.22	1.6
Ventridens	33	143	0.23	1.8

replicates necessary for bootstrap proportions to be reliable is much higher than typically used in phylogenetic studies. We attempted to bootstrap our data using the general heuristic algorithm in PAUP 3.1.1 (Swofford, 1993). Unfortunately, some bootstrap replicates generated more than 5,000 equally parsimonious trees, and with available computing power, it was not possible to complete the several hundred replicates needed to obtain reliable bootstrap values. Felsenstein (1985), however, has shown that when all characters are perfectly compatible, bootstrapping will show significant support for a group if it is defined by at least three characters. Therefore, groups defined by fewer than three characters in Figure 3 cannot be considered to have significant support. Groups defined by three or more characters might not be significant if some of the characters are homoplastic, or incorrectly polarized.

Because an explicit, morphology-based cladistic analysis of gastropod phylogeny has not yet been published (see Bieler, 1990), we could not perform a combined analysis of molecular and morphological data (see Bull *et al.*, 1993; Eernisse & Kluge, 1993). We have instead attempted to evaluate on a case-by-case basis the evidence supporting the major clades shown in Figure 3. Particular characters are referred to in the form "21>A," meaning position 21, character state A.

Archaeogastropoda

Two taxa traditionally classified as archaeogastropods were included in the analysis, Neritopsina (*Helicina*) and Pleurotomariidae (*Perotrochus*). The Neritopsina (= Neritimorpha, Neritoida) traditionally have been regarded as having strong affinities with the Caenogastropoda (Bieler, 1992). Recent classifications, however, have reversed this, with Neritopsina being placed basally to the Archaeo- and Caenogastropoda (Haszprunar, 1988a, b; Healy, 1988; Hickman, 1988) on the basis of lack of skeletal rods in the ctenidium, and on sperm morphology. Haszprunar (1988a, b) regarded similarities of Neritopsina and Caenogastropoda as convergences due to the specialized reproductive biology of Neritopsina.

The sequence data show Pleurotomariidae rather than Neritopsina to be more closely related to the other gastropods, supporting Haszprunar's morphology-based classification. Trees with branching order (Pleurotomariidae (Neritopsina (Caenogastropod Pulmonata))) are three steps longer than those with the order (Neritopsina (Pleurotomariidae (Caenogastropod (Pulmonata))). The latter tree is supported by three characters: 94>A, 136>A and 143>U. Only a partial sequence was obtained for Pleurotomariidae, and several positions (32, 42, 51, 53), may also support its derived position relative to Neri-

topsina. The position of Pleurotomariidae is discussed further under Caenogastropoda.

Neritopsina groups with the Bivalvia based on three characters (31>A, 59>A and 88>U) whereas only a single character supports grouping it with the Gastropoda (133>G). All other nucleotide positions are uninformative as to its relationships. The three characters supporting affinity with bivalves may prove plesiomorphic for mollusks when sequences from other molluscan classes are added. The sequence of Neritopsina has diverged relatively little from that of bivalves, and several places where it has changed are uniquely autapomorphic (positions 21, 26, 82, 120, 135, 144). Sequences from additional taxa such as Patellogastropoda, Cocculiniformia, and Neomphalidae might show that some of the autapomorphies of Neritopsina are actually basal synapomorphies that will unite the Gastropoda.

Caenogastropoda

Neotaenioglossa, Neogastropoda and Pleurotomariidae group together in our analysis, with Pleurotomariidae basal to the Neogastropoda. This is consistent with Ponder's (1973) hypothesis of archaeogastropod origins of the Neogastropoda, and contradicts the monophyly of Caenogastropoda. However, osphradial and spermatozoic characters have been found recently that indicate that the Neogastropoda probably evolved from the higher mesogastropods (Haszprunar, 1988a, b; Healy, 1988; Taylor & Morris, 1988). In Figure 3, sixteen characters define branches leading to the *Perotrochus* lineage. In *Perotrochus*, no data are available for ten of these because of the incomplete sequence, and two positions differ (80>U and 81>G). The single character that unites Pleurotomariidae with Neogastropoda in our analysis (124>C) is insufficient to refute morphological characters defining the Neogastropoda and is likely convergent. Addition of taxa such as Littorinoidea or Cerithioidea that would intersect the branch between Pleurotomariidae and Neogastropoda might reveal this convergence, as might completion of the partial sequence for Pleurotomariidae. Because the position of *Perotrochus* in Figure 3 is weak, it cannot be taken as contradicting the monophyly of Caenogastropoda or Neogastropoda.

Neogastropoda

Taylor and Morris (1988) concluded on the basis on morphological characters that Neogastropoda is monophyletic. If *Perotrochus* is excluded, the nucleotide sequences supports this monophyly, character 61>C being a unique synapomorphy for the Neogastropoda. Another character also supports the node (76>G) but with homoplasy elsewhere in the tree. The relationships of the three neogastropod taxa used in our study differ from those postulated by Taylor and Morris (1988) who regarded Cancellarioidea to be a possible sister group of Rachiglossa + Conoidea (i.e., Stenoglossa + Toxoglossa). In our analysis, the Muricidae and Cancellariidae group together, with Melongenidae as a sister group. Three characters support the monophyly of Muricidae and Cancellariidae: 37>U (an insertion), 58>G and 113>C, and none contradict it.

The relationship of Cancellariidae to other neogastropod taxa has been uncertain, with the group having been included by various authors in the Toxoglossa, Stenoglossa, and its own order, the Nematoglossa (see Petit & Harasewych, 1990). Most subsequent authors have followed Ponder (1973) in dividing the Neogastropoda into the corresponding superfamilies Conoidea, Muricoidea and Cancellarioidea. More recently, Kantor and Harasewych (1992) noted anatomical similarities between Cancellariidae and the stenoglossan family Volutomitridae, and suggested that a reassessment of the taxonomic rank and systematic position of these taxa was warranted. The present data support the hypothesis that the Cancellariidae comprise a highly derived group within the Stenoglossa, as reflected in such earlier classifications as those of Thiele (1929) and Wenz (1943).

Neotaenioglossa

All of the neotaenioglossans studied are rissooideans. Five characters support the monophyly of Rissooidea: 34>U, 39>U, 52>G, 93>C, 115>U. All of these are uniquely derived, except 93>C, which is convergent with *Helicina*. Within Rissooidea, *Geomelania* grouped with the Pomatiopsidae rather that the Truncatellidae, but a number of morphological characters argue that it is a truncatellid. These include truncation of the apical whorls of the shell, reduction in the number of anterior rachidian cusps, the looping mode of locomotion, and shortening of the pleuro-supraesophageal connective (Davis, 1979; Rosenberg, 1989). The single nucleotide character that appears synapomorphic for *Geomelania* and Pomatiopsidae (139>U) may prove to be plesiomorphic for rissooideans when sequences from more taxa, such Assimineidae and Hydrobiidae are added to the analysis. Within the Truncatellinae, relationships were unresolved, except that *Truncatella scalaris* and *Truncatella clathrus* grouped together as sister species. This grouping is confirmed by allozyme and morphological data (Rosenberg, 1989).

Pulmonata

Our alignment for species treated by Emberton *et al.* (1990) differs in minor details from their published alignment, and polarities of some characters have changed with added data, but these differences do not affect inferred relationships among the pulmonates. The monophyly of Pulmonata is strongly supported by seven characters and monophyly of Stylommatophora is supported by three characters. No other nodes within the pulmonates are supported by more than two characters.

Bivalvia

The sequence data support the distinction between Ambleminae and Margaritiferinae advocated by Davis and Fuller (1981), but do not reflect the distinctiveness of Anodontinae, which nests within Ambleminae (represented by *Amblema* and *Gonidea* in Figure 3). Six characters give support for grouping Anodontinae and Ambleminae apart from the Margaritiferinae, but a few of these would be plesiomorphic if *Helicina* were rerooted. Only character 133>C supports the grouping of *Anodonta* with some of the amblemines. Unpublished sequences from the 5′ terminus obtained during this study were also uninformative, containing only a single variable site, corresponding to position 23 of Emberton *et al.* (1990). Given the large body of immunological, electrophoretic, and anatomical evidence showing the distinctiveness of the Anodontinae from the Ambleminae (Davis & Fuller, 1981, Davis *et al.*, 1981), we maintain the tripartite subfamilial classification of Unionidae, with Margaritiferinae as sister group to the clade containing Ambleminae and Anodontinae. The sequences for *Unio* and *Amblema* are identical, indicating that Ambleminae may be a synonym of Unioninae, however, we refrain from changing the taxonomy until more data sets for *Unio* are available.

Sequence Variability in Mollusks

Emberton *et al.* (1990) found that 13% of sites in the D6 divergent domain of 28S rRNA were informative for stylommatophoran phylogeny, but only 1% in the D6 flanking regions were informative. They concluded that divergent domains of LrRNA would be of "some value in resolving stylommatophoran phylogeny." We have found that the D6 region is more variable in mollusks than anticipated from the results of the first study, with 23% of sites in the D6 flank and more than 70% in the divergent domain being informative at some level for molluscan phylogeny.

Variability in the bivalves and pulmonates is almost entirely in the form of nucleotide substitutions; there are only a few insertions and deletions. In the caenogastropods, the D6 loop region is subject to considerable variation in length, in addition to nucleotide substitutions.

The selection of taxa for phylogenetic analysis is extremely important because taxa display varying rates of DNA sequence evolution (slower in unionoideans, faster in rissooideans); differing rates in different regions (slower in the D6 flanks, faster in the D6 loop); and different proportions of substitutions versus insertion and deletions in various taxa. A single species or genus often is not representative of its higher taxon, as three examples show.

1.) *Helicina* has diverged from ancestral sequences more slowly than the other gastropods, and groups with the slowly evolving bivalves, perhaps because of retained plesiomorphies, as discussed above.

2.) *Margaritifera falcata* has eight autapomorphies, whereas *M. margaritifera* has none, and *Cumberlandia* has three (Figure 3). No other unionoidean had more than one autapomorphy. There were no convergent autapomorphies between *M. falcata* and *Cumberlandia*, but there easily could be between derived unionoidean species. Such convergences can be mistaken for synapomorphies, as we interpreted happened between *Anodonta* and some amblemines.

3.) *Truncatella* resembles *Mancinella* and *Progabbia* in having long D6 loop sequences; *Geomelania* and *Oncomelania* resemble *Busycon* in having short D6 loop sequences. From positions 65 to 74, *Truncatella sp.* shares six of eight nucleotides with *Progabbia*. The convergence is due in part to compensatory mutations in the D6 loop. *Truncatella sp.* has nine adenines between positions 72 and 82 some of which base-pair with six uracils between positions 58 and 70. *Progabbia* has a corresponding A/U rich region. The convergence is revealed as such by comparison to the sequences of close relatives. Thus, *Truncatella clathrus* shares no nucleotides with *T. sp.* in the region where *Progabbia* shares six, but it is identical in the D6 flanking regions.

Because species may not be representative of their higher taxa, and because long branches attract, a mixture of closely and distantly related species must be incorporated into phylogenetic analyses to minimize the chance of convergences remaining undetected.

If convergence is undetected, it is often revealed by comparison to relationships inferred from other data sets. In our analysis, comparisons to morphological phylogenies have identified several areas where the sequence data by themselves seem to be misleading. Problematic areas in Figure 3, the grouping of *Anodonta* with *Amblema*, *Helicina* with the bivalves, *Perotrochus* with the neogastropods, and *Geomelania* with *Oncomelania*, are not strongly supported by the sequence data, with the alternative trees being in each case only one or two steps longer. In the cases of *Helicina* and *Geomelania*, the addition of taxa is likely to change the polarity of those characters supporting the doubtful groupings. That is, the problem is not necessarily convergence, but that plesiomorphy has been mistaken for apomorphy. With *Perotrochus*, addition of taxa such as patellogastropods might also show that characters scored as autapomorphic are synapomorphic with basal taxa. This would reveal the convergence that groups *Perotrochus* with neogastropods. In the case of *Anodonta*, adding taxa would not help, but more variable sequences from another region might.

To date, sequence studies of mollusks have not overturned phylogenies based on morphology, but rather have helped in choosing among competing morphology-based hypotheses. Like morphological data, sequence data are subject to problems of convergence, unequal rates of evolution, and choice of taxa. Phylogenies based on sequence data alone can be misleading. Molecular and morphological data are often complementary, serving to define different nodes within a tree. Analyses and classifications must be based on all available data to maximize

the potential for detecting convergences and correctly resolving phylogenetic relationships.

ACKNOWLEDGMENTS

Supported in part by NIH grant AI 11373-TMP to George M. Davis. We thank Edward Theriot and two anonymous reviewers for helpful comments on the manuscript, Caryl Hesterman and Kathleen M. Monderewicz for technical assistance, and S. Michael Phillips for providing laboratory space at the University of Pennsylvania.

LITERATURE CITED

Albert, V. A. and B. D. Mishler. 1992. On the rationale and utility of weighting nucleotide sequence data. Cladistics 8:73–83.

Bieler, R. 1990. Haszprunar's "clado-evolutionary" classification of the Gastropoda–a critique. Malacologia 31:371–380.

Bieler, R. 1992. Gastropod phylogeny and systematics. Annual Review of Ecology and Systematics 23:311–338.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford, and P. J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Systematic Biology 42: 384–397.

Crother, B. I. and W. F. Presch. 1992. The phylogeny of xantusiid lizards: The concern for analysis in the search for a best estimate of phylogeny. Molecular Phylogenetics and Evolution 1:289–294.

Davis, G. M. 1979. The origin and evolution of the gastropod family Pomatiopsidae, with emphasis on the Mekong River Triculinae. Academy of Natural Sciences of Philadelphia, Monograph 20: vii + 120 pp.

Davis, G. M. and S. L. H. Fuller. 1981. Genetic relationships among Recent Unionacea (Bivalvia) of North America. Malacologia 20:217–253.

Davis, G. M., W. H. Heard, S. L. H. Fuller and C. Hesterman. 1981. Molecular genetics and speciation in *Elliptio* and its relationships to other taxa of North American Unionidae (Bivalvia). Biological Journal of the Linnean Society 15: 131–150.

Dixon, M. T. and D. M. Hillis. 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. Molecular Biology and Evolution 10:257–267.

Eernisse, D. J. and A. G. Kluge. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules and morphology. Molecular Biology and Evolution 10:1170–1195.

Emberton, K. C., G. S. Kuncio, G. M. Davis, S. M. Phillips, K. M. Monderewicz, Y. H. Guo. 1990. Comparison of Recent classifications of stylommatophoran land-snail families, and evaluation of large-ribosomal-RNA sequencing for their phylogenetics. Malacologia 31:327–352.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27:401–410.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Field, K. C., G. J. Olsen, D. J. Lane, S. J. Giovannoni, M. T. Ghiselin, E. C. Raff, N. R. Pace, and R. A Raff. 1988. Molecular phylogeny of the animal kingdom. Science 239: 748–753.

Ghiselin, M. T. 1988. The origin of molluscs in light of molecular evidence. Oxford Surveys of Evolutionary Biology 5:66–95.

Gutell, R. R. and G. E. Fox. 1988. A compilation of large subunit RNA sequences presented in a structural format. Nucleic Acids Research Sequences Supplement 16:175–269.

Hassouna, N., B. Michot and J.-P. Bachellerie. 1984. The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucleic Acids Research 12:3563–3583.

Haszprunar, G. 1988a. A preliminary phylogenetic analysis of the streptoneurous gastropods. Malacological Review, Supplement 4:7–16.

Haszprunar, G. 1988b. On the origin and evolution of the major gastropod groups, with special reference to the Streptoneura. Journal of Molluscan Studies 54:367–441.

Healy, J. M. 1988. Sperm morphology and its systematic importance in the Gastropoda. Malacological Review, Supplement 4:251–266.

Hedges, S. B. 1992. The number of replications needed for accurate estimation of the bootstrap *P* value in phylogenetic studies. Molecular Biology and Evolution 9:366–369.

Hedges, S. B., K. D. Moberg and L. R. Maxson. 1990. Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. Molecular Biology and Evolution 7:607–633.

Hickman, C. S. 1988. Archaeogastropod evolution, phylogeny and systematics: a re-evaluation. Malacological Review, Supplement 4:17–34.

Hillis, D. M. and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42:182–192.

Hillis, D. M. and M. T. Dixon. 1991. Ribosomal DNA: Molecular evolution and phylogenetic inference. Quarterly Review of Biology 66:411–453.

Kantor, Y. I. and M. G. Harasewych. 1992. Morphology of the digestive system of *Volutomitra alaskana* Dall, 1902 (Gastropoda, Pectinibranchia, Volutomitridae), with notes on the possible mechanism of feeding. Ruthenica 2:45–53.

Maddison, W. P. and D. R. Maddison. 1992. MacClade Version 3. Sinaeur Associates: Sunderland, Massachusetts. xi + 398 pp.

Mishler, B. D., K. Bremer, C. J. Humphries and S. P. Churchill. 1988. The use of nucleic acid sequence data in phylogenetic reconstruction. Taxon 37:391–395.

Petit, R. E. and M. G. Harasewych. 1990. Catalogue of the superfamily Cancellarioidea Forbes and Hanley, 1851 (Gastropoda: Prosobranchia). The Nautilus, Supplement 1, 69 pp.

Ponder, W. F. 1973. The origin and evolution of the Neogastropoda. Malacologia 12:295–338.

Ponder, W. F. and A. Warén. 1988. Classification of the Caenogastropoda and Heterostropha—A list of family-group names and higher taxa. Malacological Review, Supplement 4:288–328.

Rosenberg, G. 1989. Phylogeny and evolution of terrestriality of the Atlantic Truncatellidae (Prosobranchia: Gastropoda: Mollusca). Ph.D. thesis, Harvard University. x + 296 pp.

Smith, A. B. 1989. RNA sequence data in phylogenetic reconstruction: testing the limits of its resolution. Cladistics 5:321–344.

Sogin, M. L., H. J. Elwood, and J. H. Gunderson. 1986. Evolutionary diversity of eukaryotic small-subunit rRNA genes. Proceedings of the National Academy of Sciences, USA 83:1383–1387.

Steele, K. P., K. E. Holsinger, R. K. Jansen, and D. W. Taylor. 1988. Phylogenetic relationships in green plants—A comment on the use of 5S ribosomal RNA sequences by Bremer et al. Taxon 37:135–138.

Swofford, D. L. 1993. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1.1. Computer program distributed by the Illinois Natural History Survey.

Swofford, D. L. and G. J. Olsen. 1990. Phylogenetic reconstruction. Pp. 411–501 *in* D. M. Hillis and C. Moritz (eds.), Molecular Systematics. Sinauer Associates: Sunderland, Massachusetts.

Tautz, D., J. M. Hancock, D. A. Webb, C. Tautz and G. A. Dover. 1988. Complete sequences of the rRNA genes of *Drosophila melanogaster*. Molecular Biology and Evolution 5:366–376.

Taylor, J. D. and N. J. Morris. 1988. Relationships of neogastropods. Malacological Review, Supplement 4:167–179.

Thiele, J. 1929. Handbuch der Systematischen Weichtierkunde. 1(1). Loricata./Gastropoda. I:Prosobranchia (Vorderkiemer). Jena: 376 pp.

Tillier, S., M. Masselot, H. Philippe and A. Tillier. 1992. Phylogénie moléculaire des Gastropoda (Mollusca) fondeé sur le séquençage partiel de l'ARN ribosomique 28 S. Comptes Rendus de l'Académie des Sciences 314:79–85.

Wenz, W. 1943. Gastropoda. Teil I: Allgemeiner Teil und Prosobranchia. Pp. 1201–1506 (1943) *in* O. H. Schindewolf, ed., Handbuch der Paläozoologie, vol. 6. Berlin.

Wheeler, W. C. 1990. Combinatorial weights in phylogenetic analysis: a statistical parsimony procedure. Cladistics 6:269–275.

Wheeler, W. and D. Gladstein. 1993. MALIGN. Version 1.73 1/4/93. Published by the authors. 28 pp.

Wheeler, W. C. and R. L. Honeycutt. 1988. Paired sequence difference in ribosomal RNA's: evolutionary and phylogenetic implications. Molecular Biology and Evolution 5:90–96.

Woese, C. R. and G. J. Olsen. 1986. Archaebacterial phylogeny: perspectives on the urkingdoms. Systematic and Applied Microbiology 7:161–177.

Monophyly of Major Gastropod Taxa Tested from Partial 28S rRNA Sequences, with Emphasis on Euthyneura and Hot-Vent Limpets Peltospiroidea

Simon Tillier[1]
Monique Masselot[2]
Jean Guerdoux[2]
Annie Tillier[1]

[1] Laboratoire de Malacologie (CNRS URA 699 and GDR 1005)
Muséum national d'Histoire naturelle
55, rue Buffon,
F–75005 Paris, France
[2] Centre de Génétique moléculaire (CNRS UPR 2420)
CNRS
Avenue de la Terrasse
F–91198 Gif-sur-Yvette, France

ABSTRACT

In a sample of 24 species representing six higher Gastropoda taxa (Patellogastropoda, Vetigastropoda, Neomphalina, Caenogastropoda, Opisthobranchia, Pulmonata) plus outgroups, each represented by a sequence of ca. 300 nucleotides from the 5′ end of their 28S rRNA, repeated jackknives and bootstraps using either variable numbers of nucleotides or all the possible monospecific combinations of higher taxa support: (a) monophyly of the Vetigastropoda and position of the Fissurelloidea as the sister-group of the Pleurotomarioidea and Trochoidea; (b) monophyly of the Caenogastropoda including Stenoglossa, but not monophyly of the latter, which, however, is neither supported nor rejected; (c) monophyly of the Anaspidea and Pulmonata; (d) monophyly of the Neomphalina and Apogastropoda, with the former the sister-group of the latter. The monophyly of the Euthyneura is but weakly supported, as is the monophyly of the three non-Anaspidean Opisthobranchia included in the sample. The monophyly of the Patellogastropoda and Vetigastropoda is more likely than their paraphyly in this data set, but may be artefactual.

Key-words: RNA, Gastropoda, phylogeny, bootstrap, jackknife, hydrothermal-vent limpets.

INTRODUCTION

After decades of relative stagnation, hypotheses on Gastropod phylogeny at various levels have recently been raised in greater number than ever before: for example, the definition and relationships of various groups of prosobranchs (e.g. Ponder (ed.), 1988; Haszprunar, 1988a, c) and of pulmonates (e.g. Schileyko, 1979; Tillier, 1989; Emberton et al., 1990) have been revised and are still under discussion. In addition, the discovery of new, primitive-like Gastropoda from such extraordinary environments as deep-sea hydrothermal vents has brought still more interest to attempts to analyze ancient Gastropod radiations (e.g. McLean, 1981), and their relationships are still subject to discussion (e.g. McLean, 1990a, b).

In his phylogenetic classification of the Gastropoda, Haszprunar (1988a, c) proposed the successive emergences of thirteen higher taxa, as shown modified in figure 1:

(a) Docoglossa = Patellogastropoda, of which the distinctiveness is recognized by all recent authors (see Lindberg, 1988);
(b) "hot-vent c" = Neolepetopsidae McLean, 1990, hot-vent limpets later included in the Patellogastropoda (McLean, 1990a);
(c) Cocculiniformia, of which most are deep-sea limpets (Haszprunar, 1988b);
(d) Neritimorpha = Neritopsina, which include marine, freshwater and terrestrial taxa;
(e) Neomphalina = "hot-vent a" = "tapersnout";
(f) Vetigastropoda, which correspond to most of the classical "archaeogastropods";
(g) Seguenziina (Seguenziidae);
(h) Architaenioglossa, which include the land and freshwater Cyclophoroidea and Ampullarioidea, which are included with all the former taxa in the grade Archaeogastropoda by Haszprunar (1988c) but are considered as Caenogastropoda by Ponder & Warén (1988);
(i) Caenogastropoda, which include most of the classical "mesogastropods" and the supposedly monophyletic Stenoglossa = Neogastropoda;
(j) Campanilimorpha (Campanilidae), considered by Ponder & Warén (1988) as members of the Caenogastropoda; a position finally admitted by Haszprunar (1992);
(k) Ectobranchia = Valvatoidea;
(l) Allogastropoda, a paraphyletic (?) taxon including

several families of small snails, marine or rarely freshwater, with a hyperstrophic protoconch in addition to several anatomical synapomorphies;

(m) Euthyneura, including the Opisthobranchia (mainly marine) and the Pulmonata (mainly freshwater and terrestrial).

All the taxa included in the Valvatoidea, Allogastropoda and Euthyneura may be grouped in the Heterostropha (Ponder & Warén, 1988).

Even at such high taxonomic levels, many points are still under discussion. In this context, molecular systematics, and in particular sequences of nucleic acids are useful tools to bring new insights to phylogenetic relationships and to test hypotheses currently under discussion. In former papers, Emberton et al. (1990) and Tillier et al. (1992) have tested the use of 28S rRNA to elucidate gastropod phylogeny and the latter have shown that the 5′ extremity of 28S rRNA is suitable for partial resolution of gastropod phylogeny at high taxonomic levels.

For any sequencing effort, a choice must necessarily be made at some point between obtaining short sequences from many taxa, or obtaining long sequences from a few taxa. The second option allows use of many characters, and retains many characters per taxon even when all regions of even slightly doubtful homology are eliminated. However, this option is not necessarily the better one since species sampling has itself a strong influence on the reliability of the groupings found (Lecointre *et al.*, 1993). Here we use a data set corresponding to the first case, which necessitates either the use of sites from variable regions in order to keep the ratio of characters per taxon sufficiently high (here between 3 and 4), with resulting low reliability due to high homoplasy; or using statistical methods which allow estimation of the reliability of various groupings with and without variable regions. The latter is done here using the methods proposed by Lecointre *et al.* (1993 and in press). As shown further, these latter methods may prove powerful for testing the consistency of the trees under discussion with a given set. However, one should never forget that they are not doing more, and in particular one should not confuse the reliability of the trees as a representation of a given data set, which is discussed further, and their reliability as a representation of the history of life, which certainly cannot be estimated by internal statistical tests.

In the present paper, we address several of the points under discussion in gastropod phylogeny using homologous sequences from 24 species listed in table 1 and figure 1. These species were chosen from those available to represent the diversity of the monophyletic taxa listed above: Patellogastropoda, Vetigastropoda, Neomphalina, Caenogastropoda including Stenoglossa, Opisthobranchia and Euthyneura. Both new data and application of methods more sophisticated than those used formerly, lead us to discuss two types of questions: some already contested points, like the para- or monophyly of the Archeogastropoda represented in our sample or the monophyly of Neomphalina, Caenogastropoda and Euthyneura; or to raise some seemingly formerly neglected points, like the monophyly and relationships of the Opisthobranchia and Pulmonata.

MATERIALS AND METHODS

Materials

Samples from 24 species were used (tables 1 and 2, figure 1): six Pulmonata (two Archaeopulmonata, one Basommatophora, and one Elasmognatha in addition to two Stylommatophora); four Opisthobranchia (one of each of the Cephalaspidea, Notaspidea, Anaspidea and Nudibranchia); six Caenogastropoda; one Neomphalina (the hot-vent Peltospiridae limpet *Rhynchopelta concentrica* McLean, 1989); four Vetigastropoda; one Patellogastropoda; plus one Polyplacophora and one Bivalvia as outgroups. The sequences correspond to the 5′ end of the 28S rRNA, starting from position 72 in the sequence of *Mus musculus* in Hassouna et al. (1984), and include 214 to 222 nucleotides (alignment with the mouse sequence, table 1; positions homologous to positions 1–71 in the mouse have not been used because we miss them in *Buccinum*, *Pomatias*, and *Haliotis*).

Total RNA was extracted from fresh or frozen tissues by the guanidium thiocyanate-phenol-chloroform method (Chomczynski & Sacchi, 1987), sometimes using RNAzol (Bioprobe) for extraction, and sequenced directly using the method of Qu et al. (1983), with either ^{32}P or ^{35}S as markers. The sequence of the probe is 5′-AGGTAGATTCCGATTTATATGGCCGT-3′. Every sequence was repeated from three to nine times and each repetition was read at least two times by two different readers (M.M. and A.T.). When necessary, sequencing was repeated using deazaGTP (Barr et al., 1986).

Sequences were registered and handled, from database to tree analysis, using the MUST package (Philippe, 1992). The alignment was realized entirely by eye. In cases of ambiguity in the position of a gap, gaps have been preferably clustered between sequences, with the underlying assumption that the occurrence of deletions or insertions is more likely at homologous sites, and that substitutions are more likely than independant deletions or insertions in the ambiguous region (this maximizes the homology of the gaps and minimizes the number of independant deletions/insertions); and gaps have been preferably clustered inside a given sequence, considering that elongation of a gap is more likely than occurrence of a new independant gap in the immediate neighbouring region of the molecule.

In all analyses, only characters informative for parsimony (i.e. exhibiting at least two states occurring in more than one taxon) were used: 93 sites in the whole sequences, 65 sites for 24 species when the more variable regions overlined in table 1 are removed. Relative rate tests (Wilson et al., 1977) were done both with and without the variable regions overlined in table 1; all kinds of differences (transitions, transversions and gaps) have been used with equal weights for this test as well as in the various calculations.

Table 1. The raw data set: 28S r-RNA sequences oriented from 5′ to 3′, starting from position 72 of the mouse (included for comparison, last line). Variable regions, removed from all analyses except BP/number of sites diagrams, are overlined. Identities with the sequence of *Helix aspersa* (upper line) are denoted by hyphens, deletions by stars and undetermined nucleotides by the standard code N. Numbers in the right row indicate the actual number of nucleotides.

Helix aspersa	AAGGAUUCCC	UCAG*UAACG	GCGAGUGAAG	CGGUAAUAGC	CCAGCACC**	GAAUCCCUCA	***GUGUCAU	*GCUG*ACGG	GAA*CUGUGG	81
Placostylus fibratus	-------U--	C---*-----	----------	---N--G---	--------**	----------	***-------	*----*----	N--*------	81
Phytia myosotis	N------U--	C---*-N---	----------	---N--G---	--------**	----------	***-------	*----*----	---*------	81
Siphonaria algesirae	N------U--	C---*-----	----------	---A-UG---	--------**	----------	***----G-C	*----*----	---*------	81
Lymnaea stagnalis	-------U--	C---*-----	----------	---G--G---	--------**	-------C--	***------C	*----*G---	---*------	81
Succinea putris	-------U--	C---*-----	----------	------A---	--------**	-------C--	***------C	*----*G---	---*------	81
Aplysia depilans	--C----U--	CU--*-----	----------	---A--G---	--------**	-------C--	***----A--	*----*G---	---*------	81
Archidoris tuberculata	C---------	C---*-----	----------	------G---	--------**	-------C-G	***-C*C-UC	C-UC-*G---	U--*------	81
Actaeon tornatilis	U------U--	C---*-----	----------	---G--G---	--------**	---------G	***-C---UC	*-UC-*----	---*------	81
Berthella plumula	N---------	C---*-----	--N-------	---N--A---	--------**	-------C-G	***-C-A--C	C-UC-*GN--	N--*------	82
Ocenebra erinacea	----------	----*N----	N---------	---G-CC---	--------**	-------C--	***-CA--UC	*----*G---	---*------	81
Nucella lapillus	----------	----*-----	----------	---G-CC---	--------**	-------C--	***-CA--UC	*----*G---	---*------	81
Buccinum undatum	N---------	----*-N---	----------	---G-UC---	--------**	-------C--	***-CA-GUC	*----*G---	---*------	81
Calyptraea chinensis	----------	----*-----	----------	-----UC---	--------**	-------C--	***-CA--UC	*----*G---	-CG*A-----	81
Pomatias elegans	C---------	----*GN---	----------	---N--G---	--------**	-------C--	***--CGACG	*----*GU--	U--*--A---	81
Littorina littorea	NN--------	----*-----	-N--------	---G-UC---	--------**	-------C--	***--C-UUG	*----*G---	N--*------	81
Rhynchopelta concentrica	----------	----*-----	----------	---G--G---	--------**	-------GAG	****C---GC	*----*C---	---*--N---	80
Monodonta lineata	----------	-U--U-----	----------	------G---	--------**	-------CUG	***--UCUGC	*---A*G---	---*U-----	82
Calliostoma zizyphinum	----------	----*-----	----------	---G--G---	--------**	-------CUG	***--UCUGC	*--CA*G---	--G*U-N---	81
Haliotis tuberculata	N---------	--NN*-N---	A---------	---N--C---	--------**	-------CUG	***--UCUGC	*N-C-*G---	---*--N---	81
Diodora graeca	----------	----*-----	**--------	---G--C---	--------**	-------C--	***-G-GA-A	CU---*G---	-G-A------	81
Patella vulgata	CG--------	----*--G--	A----C----	---N--G---	---U-G--GA	-------C-G	CUC-GAC-UC	G--C-UU---	---*U-----	88
Acanthochitona fascicular	N---------	GU--*-----	*N--------	---G--G---	-----G--**	-------A-G	****GC--G-	A-UC-*CU--	---*------	80
Mytilus edulis	U---------	CU--*---U-	----A-----	---G--G---	U-------**	-------G--	***-CC-UGC	GCUGC*-G--	A-**------	81
Mus musculus	C---------	----*-----	---------C	A--G--G---	-----G--**	-------CGC	CGC-C---GC	G--GU**G--	A--UG--GC-	85

Helix aspersa	UGUGUGGGAC	GCCACCAGUC	GCA**UCAGA	GG****GCGC	CGAAGUCCUC	CUGAUCGGGG	CUUCACCCAG	*AGCGGGUGU	AAGGCCUUU*	163
Placostylus fibratus	----------	----------	---**--G--	--****-U--	----------	----------	----------	*------N--	---------*	163
Phytia myosotis	----------	----------	---**----C	--****----	----------	-------A--	----------	*--U------	---------*	163
Siphonaria algesirae	--------N-	----------	-U-**----C	--****A-A-	----------	-------A--	----------	*------N--	---------*	163
Lymnaea stagnalis	----------	----------	---**----C	--****--A-	----------	-------A--	----------	*---------	---------*	163
Succinea putris	----------	----------	--G**C-G--	--****-U--	----------	----------	----------	*---------	---------*	163
Aplysia depilans	----------	----------	--G**CGU-C	-A****--A-	----------	-------A--	----------	*---------	---------*	163
Archidoris tuberculata	----C-----	--------C-	--G**CG--C	--****----	----------	-------A--	-C--U----C	*G-U------	---------*	163
Actaeon tornatilis	----------	----A-U---	-AG**CGC-C	--***G--C-	-C--------	----------	---UCU---U	*---------	---------*	164
Berthella plumula	---A----N-	--N-A---C-	UUG**C-C-G	AC***G--A-	AC--------	----NA-A--	-CCUU---G*	*U--------	G--------*	164
Ocenebra erinacea	---A------	----A-U---	-UC**--CCC	--****UG--	----------	----------	-C--U-----	*---------	C--------*	163
Nucella lapillus	---A------	----A-U---	-UC*UC-GCC	--****UG--	----------	----------	-C--U-----	*---------	C--------*	164
Buccinum undatum	---A------	-U----U---	-A-*UG--CC	--****UNU-	----------	----------	-C--U-----	*--U------	C--------*	164
Calyptraea chinensis	---A------	----A-U---	-UC*CG--CC	--****UG--	----------	----------	-C--U-----	*--U------	C--------*	164
Pomatias elegans	---A--A---	-U--A-U---	-UC*CGAUCC	--****UG--	----------	----NN--N-	-C--------	*--U------	N----N--N*	164
Littorina littorea	---A------	----A-U---	-UC*CGACCG	--*****GA-	----------	----------	-C--U----C	*---------	C--------*	163
Rhynchopelta concentrica	---UG-----	----A-C---	-AC*GGGUCC	UC*GUU----	-C--------	----------	-C--U----C	*G-----N--	C--------*	166
Monodonta lineata	---U----C*	*G-UAUC-G-	--U*G-GUCC	CU****-U--	-C-------U	----------	-C-UU----U	*--U------	C--------*	163
Calliostoma zizyphinum	---U-----G	*--GGUC-G-	--C*A-GUCC	CU****-U--	-C-------U	----------	-C-UU----U	*--U------	C--------*	163
Haliotis tuberculata	-----A-CGA	*N-GUUC-G-	AUC*UA-GC-	CU*CCACU--	-C--------	-------C--	-***U-----	*---------	C--------*	163
Diodora graeca	---U------	*G--UUC-G-	-GU*CG-UCU	CU****----	-C-------U	----------	-C-UU-----	G--U------	C--CG----*	164
Patella vulgata	CK-AA---U-	*-G--UC---	AGGCG--U--	CC**GUC--A	AC------CG	-G-GA-A--U	G-G-U-----	*--U------	G--------A	174
Acanthochitona fascicular	C--A-A----	AG-CU-U--G	C-G**CGUUC	--****-U--	-CG-------	----------	-C-UU-----	*---------	U------C-*	162
Mytilus edulis	---U------	-U-UAUU-G-	--G***AU-U	CC**GG-U--	-U-G------	----------	-C--U-----	*---------	C--------*	164
Mus musculus	-AC-GAA---	*----UCCC-	-GC*GC-GCU	C-UGGG-G--	-C-------U	-------A--	-CCAG---GU	*G-AC-----	G-----GG-*	171

Table 1. Continued.

Helix aspersa	*GCGGGUGC* CUCUCUGUGC GGCCGC*GAG ***CGUCUCA GGAGUCGGGU UGUUUGG	214
Placostylus fibratus	*--A-----* U----C---- ------*--- ***------- ---------- -------	214
Phytia myosotis	*--A-----* ---G------ ---*--*--- ***------- ---------- -------	213
Siphonaria algesirae	*--A-----* --G------- ------*N-- ***------U -----Y---- -------	214
Lymnaea stagnalis	*--------* UC-G------ -----A*--- ***------N ---------- -------	214
Succinea putris	*--U--CA-* -----C-C-- ------*--- ***------N ---------- -------	214
Aplysia depilans	*--U-----* *-UG---C-- ------*--- ***------- ---------- ------A	213
Archidoris tuberculata	*--U-CG--* -GGCGC-C-- ------*--- ***----C-- --------N- -------	214
Actaeon tornatilis	**----A--* *CNNGC--U- -C-*--*--- ***----C-- ---------- -------	212
Berthella plumula	*N---*---* -CG--G-CU- ---N-U*--- ***------C ---------- ------A	214
Ocenebra erinacea	*A-U-*C--* *-GG-G-GU- ---U--*--- ***------C ---------- -------	212
Nucella lapillus	*A-U--CUG* --GG-G-GU- ---U--*--- ***------C ---------- -------	215
Buccinum undatum	*A-A--CCG* --GGUGCGU- ---U--*--- ***------C ---------- -------	215
Calyptraea chinensis	*A-U--CUG* --GGUC-GU- ---U--*N-- ***----C-C ---------- -------	215
Pomatias elegans	*A-CA-GCNG --GGUC-GU- ---U--*--- ***------C --N---U--- -------	216
Littorina littorea	*A-U-----* --GGUC-GU- ---U--*--- ***------U C--------- -------	214
Rhynchopelta concentrica	*AGC--C--* GGGA--CGU- ---UC-*--- ***----C-- A--------- ------A	217
Monodonta lineata	*AGC--CA-* GGGA--AG-- -U--UU*--- ***----CUU ---------- -------	214
Calliostoma zizyphinum	*AGC--CA-* *GGA--AG-- -U--UUC--- ***----CUC -----N---- -------	214
Haliotis tuberculata	*NGC--CA-* GGGA-GAGA- -U*-UU*-C- **G----CUC ---------- -------	214
Diodora graeca	*-GC--C--* GGGAGCU-C- -U--CUC--- ***----CUC ---------- -------	216
Patella vulgata	GUUC---CU* GGG---UCUG CC-G--GC-- UCGGUA-CUG C------A-- --C----	230
Acanthochitona fascicular	*A-A--CAU* --GA-GCG-U A--UCU*--- ***----CUC A--------- -------	213
Mytilus edulis	*A-C--CA-* --GG-GUC-U -C-UCA*--- ***----CUU ---------- ------A	215
Mus musculus	*AGC--CC-* -CGG-GCGC- --G-U-*-G- ***UC-UC-C ---------- --C----	222

Both parsimony and distance methods were used. For parsimony, Hennig86 (Farris, 1988) was used to build the initial trees, and DNABOOT of the PHYLIP package (Felsenstein, 1990) was used for bootstrap. However most of the trees which are discussed here, and result from numerous jackknives and bootstraps (Felsenstein, 1985, 1988) detailed further, were obtained with the neighbor joining algorithm (Saitou & Nei, 1987). This option was taken not to favour distance methods, but simply because obtaining the millions of trees analysed further would have required much more CPU time than was available (most analyses have been performed on a PC 386 (25 MHz) and a PC 486 DX (33 MHz): even if one estimates that one run of Hennig86, with options m* and bb*, takes 10 seconds, one million runs would take about four months). Repeated jackknives of sites followed by bootstrap estimates and the corresponding diagrams have been obtained by the JACKBOOT, NJBOOT and COMP BOO programs of the MUST package. Combinations of species from higher taxa followed by bootstraps and the corresponding histograms have been obtained by the programs JACKMONO, NJBOOT and MONO HIS of the same package (Philippe, 1992; Lecointre *et al.*, 1993 and in press).

Testing the Reliability of the Trees

As noticed by many authors and discussed in detail by Lecointre *et al.* (1993 and in press), whose procedures are followed here, although for partly different purposes, two principal factors determine the reliability of trees, given accepted homologies and a method of tree construction: the number of characters, here approximately proportional to sequence length, and taxonomic sampling.

Sequence Length:

A too short sequence length is generally invoked when a node considered as true appears unreliable, and when its bootstrap proportion is low. Although unavoidable in our present state of knowledge of molecular evolution, this argument should not be used without discussion since it is unfalsifiable: whatever sequence length is used, one can always invoke its shortness when expected results are not met or prove unreliable for a given data set. Bootstrapping, i.e. character sampling with replacement, is usually performed to estimate the reliability of trees (Felsenstein, 1985): a tree is constructed from each of a high number of replicate samples (here 1000), and the number of times each grouping has been found is totaled. A consensus tree is built using all majority nodes, i.e. keeping the node which is the most abundant of several conflicting nodes. In the resulting tree a grouping, i.e. a node, is considered reliable when it has been found in more than 90, or even better, 95% of the replicates; this percentage is the boostrap proportion (BP) of the node.

Lecointre et al. (in press) have proposed an empirical test for determining whether the reliability of any given node might be increased by adding a definite number of characters (here nucleotides). This test consists of: (1)

Table 2. Material studied.

Pulmonata

Helix aspersa Müller, 1774 (Stylommatophora). Around Besançon, France. 1989.
Placostylus fibratus (Martyn, 1784) (Stylommatophora). Ile des Pins, New Caledonia. Aillaud! 1990.
Succinea putris (Linné, 1758) (?Stylommatophora?, Elasmognatha). Chamarande, Essonne, France. Masselot! VI.1991.
Phytia myosotis (Draparnaud, 1801) (Archaeopulmonata). La Ville Abel, Ille et Vilaine, France. Tillier! I.1992.
Siphonaria algesirae Quoy & Gaimard, 1833 (Archaeopulmonata). Malaga, Spain. Gofas! 1991.
Lymnaea stagnalis (Linné, 1758) (Basommatophora). Gif sur Yvette, Yvelines, France. Masselot! 1990.

Opisthobranchia

Aplysia depilans Gmelin, 1791 (Anaspidea). Roscoff, Finistère, France. III.1988 and III.1992.
Archidoris tuberculata (Cuvier, 1804) = *Archidoris pseudoargus* (Rapp, 1827) (Nudibranchia). Roscoff, Finistère, France. III. 1988.
Acteon tornatilis (Linné, 1758) (Cephalaspidea). Roscoff, Finistère, France. III.1992.
Berthella plumula (Montagu, 1803) (Notaspidea). Roscoff, Finistère, France. III.1992.

Caenogastropoda

Ocenebra erinacea (Linné, 1758) (Stenoglossa, Muricidae). Roscoff, Finistère, France. III.1988.
Nucella lapillus (Linné, 1758) (Stenoglossa, Muricidae). Roscoff, Finistère, France. III.1988.
Buccinum undatum Linné, 1758 (Stenoglossa, Buccinidae). France. 1990.
Calyptraea chinensis (Linné, 1758) (Calyptraeoidea). Roscoff, Finistère, France. III.1988.
Pomatias elegans (Müller, 1774) (Littorinoidea). Between Etampes and Pierrefite, Essonne, France, Masselot! VI.1992.
Littorina littorea (Linné, 1758) (Littorinoidea). Roscoff, Finistère, France. III.1988.

Neomphalina

Rhynchopelta concentrica McLean, 1988. HERO 91, Site Justinoir, 12°48′88″N, 103°56′50″S, 2630 m. 23.X1991.

Vetigastropoda

Monodonta lineata (Da Costa, 1778) (Trochoidea). Roscoff, Finistère, France. III.1988.
Calliostoma zizyphinum (Linné, 1758) (Trochoidea). Roscoff, Finistère, France, III.1988.
Haliotis tuberculata Lamarck, 1822 (Pleurotomarioidea). Roscoff, Finistère, France. III.1988.
Diodora graeca (Linné, 1758) (Fissurelloidea). Roscoff, Finistère, France. III.1988.

Patellogastropoda

Patella vulgata Linné, 1758. Roscoff, Finistère, France. III.1988.

Polyplacophora

Acanthochitona fascicularis (Linné, 1767). Roscoff, Finistère, France. III.1988.

Bivalvia

Mytilus edulis Linné, 1758. France.

jackknife, i.e. sample without replacement, of sites in various numbers from 0 to n, n being the total number of informative sites and p the number of samplings (p=4 if jackknife is performed for 10, 20, 30 and 40 sites); (2) repeat each jackknife (for 10, 20, . . ., n–10 sites) q times; (3) bootstrap the sites, i.e. sample with replacement, for each of the pq jackknife replicates, and calculate trees for 1000 bootstrap replicates. Finally bootstrap proportions (BP) are extracted for every node having at least one BP higher than 40% occurring in the resulting pq consensus trees, for each of the p values. For each node these BP may be plotted against the number p of sites jackknifed (10, 20, 30, etc..).

From such diagrams Lecointre et al. (in press) distinguished empirically four types of nodes: those for which BP increase in value and decrease in variation in such a way that the additional sequencing necessary to reach BP higher than 90% may be estimated (e.g. figure 16); "promising" nodes, which may be interpreted as potentially entering into the first category if number of characters was increased (e.g. figure 6); those for which BP decreases in both values and variation, in such a way that one can estimate that those nodes will disappear if characters are added; and finally those nodes for which the diagram is uninformative or slightly informative (figures 7, 18), which implies that very long additional sequences are necessary to raise the BP (i.e. as far as they are "true").

In this work, p=8 and q=70: the estimates discussed further result from 560 consensuses of 1000 24 taxa—trees. This number, in addition to the 5 million trees of six or seven taxa detailed further, explains why the fast neighbor joining algorithm is used in the repetitive bootstrap procedures (NJBOOT, Philippe, 1992) rather than the very much slower DNABOOT (Felsenstein, 1990), which, to our knowledge, is the only available bootstrap program based upon parsimony for PC. All diagrams showing variation in BP as a function of the number of jackknived sites—but none of the others—have been built using the variable regions. For all points under discussion,

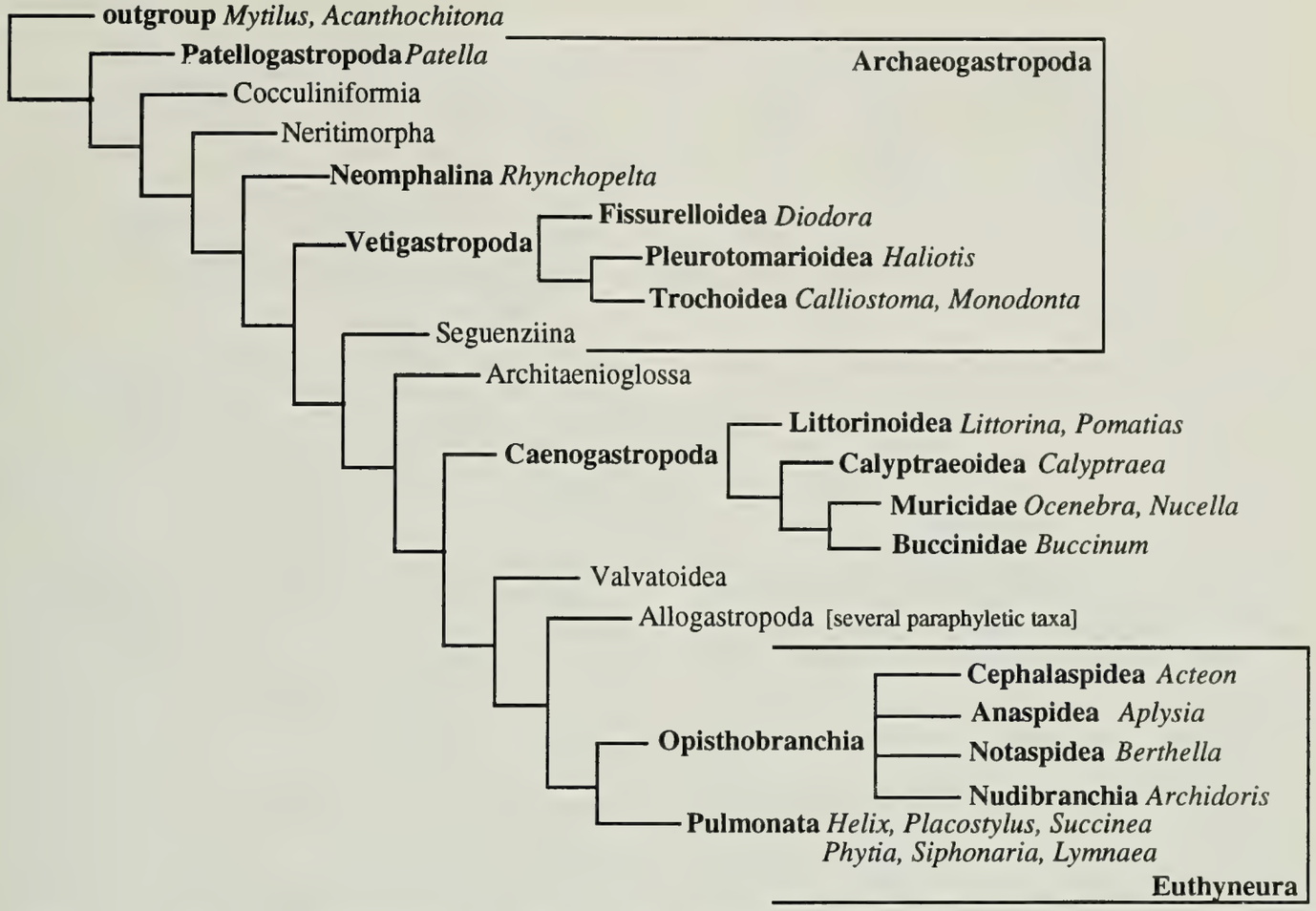


Figure 1. Positions of the species sampled, indicated by their genus name in italics, in a phylogenetic tree modified from Haszprunar's (1988c) tree. Higher taxa represented in the sample in bold. Compare with figure 29.

the diagrams made without the hypervariable regions are similar in shape to those shown here, but this shape is truncated for a value of sixty sites.

Taxonomic Sampling:

As mentioned above and shown further, minor modifications in species composition of a sample may have important effects on tree topology. These effects may be analyzed, as in Lecointre *et al.* (1993 and in press), by repeated jackknives of species followed by bootstraps. However, the problem raised in the present paper is not so much to study the stability of the nodes as to test the monophyly and relationships of pre-defined higher taxa. Another procedure proposed by Lecointre et al. (JACKMONO) is used here: it consists of establishing all possible combinations of species including a single species from each higher taxon, followed by bootstrap of sites for each combination (NJBOOT). This not only allows the analysis of the impact of any single species on higher taxa relationships, but also diminishes the possible bias due to differences in representative sampling of higher taxa under study. A histogram of BP is built for every majority node of which BP is higher than 1% (MONO HIS).

Such diagrams (e.g. figs 9, 10), not only show how BP of any node are distributed when species sample is modified, but may also be used to display jointly the distribution of all BP values for a node and the distribution of the values of the same node when it includes any given species choosen by the user: this is an efficient way to: (a) measure the impact of any given species on reliability of any node; (b) test the monophyly of any group of higher taxa, as discussed further.

Here seven higher taxa are represented: outgroup (two species), Patellogastropoda (one species), Vetigastropoda (four species), Peltospiroidea (Neomphalina, one species), Caenogastropoda (six species), Opisthobranchia (four species), and Pulmonata (six species). The total number of combinations of one species per higher taxon is thus 1152, for each of which 1000 bootstrap replicates have been calculated with the neighbor joining algorithm. Two taxa, the Patellogastropoda and Peltospiroidea, are monospecific in this study. The latter is one of the taxa under discussion; this is why the procedure has been repeated without *Patella*; it has also been repeated with and without the hypervariable regions overlined in table 1: the discussions presented further result from building approximately 5 million trees including six or seven taxa and 65 to 93 characters. All histograms pre-

Table 3. Relative rate test using either *Acanthochitona* or *Mytilus* as an outgroup. Distances in % differences (sites informative for parsimony), variable regions overlined in table 1 included.

	Helix	Placo	Phyti	Sipho	Lymna	Succi	Aplys	Archi	Actae	Berth	Ocene
Acanthochitona	65.52	65.88	67.44	72.41	67.44	65.12	65.12	63.22	58.82	64.63	58.82
Mytilus	65.56	63.64	65.91	67.42	64.04	62.92	61.80	63.33	58.62	63.10	60.23
	Nucel	Bucci	Calyp	Pomat	Litto	Rhync	Monod	Calli	Halio	Diodo	Patel
Acanthochitona	59.09	64.37	65.91	63.53	67.82	54.55	58.62	59.09	62.07	63.22	69.23
Mytilus	61.54	62.92	65.93	62.50	62.92	55.56	53.33	57.14	64.37	61.54	68.48

sented have been obtained from the data set without the variable regions overlined in table 1.

RESULTS

Relative Rate Test

In order to obtain reliable molecular phylogenetic patterns, molecular rates of evolution of included taxa must be as close as possible (Wilson et al., 1977; Felsenstein, 1978); this is tested by comparing the molecular distances of the various ingroups with the outgroups, here *Acanthochitona* and *Mytilus*, and has been done with and without the variable regions overlined table 1 (tables 3 and 4). Differences in evolutionary rates are not significant overall. In all cases: (a) *Siphonaria* and *Patella* exhibit a high relative rate of molecular evolution, the differential with other taxa being more accentuated for *Patella* when variable regions are removed; (b) *Rhynchopelta* and *Acteon* have a lower rate of substitution than other taxa in the sample; (c) when the mean rate is calculated for each of the six higher Gastropoda taxa represented, the rates are similar overall except for the faster Patellogastropoda and the slower Neomphalina—each having a single representative in the sample; the Vetigastropoda have a slightly slower mean rate than the other higher taxa represented by more than one species, whereas in Pulmonata the mean rate seems slightly higher.

Reliability of the Data: Some Ways to Obtain the Desired Tree

Without the variable regions overlined in table 1, a consensus of the 13 shortest trees found using 24 species (figure 2) shows monophyly of higher level gastropod taxa: Vetigastropoda, Caenogastropoda, and Euthyneura. However, *Succinea* emerges between *Archidoris* and *Aplysia*, i.e. between representatives of two opisthobranch orders; while *Patella* is found to be the sister-taxon of the Euthyneura. These unexpected results raise doubt as to the position of *Rhynchopelta* as the sister-taxon of the Apogastropoda (Euthyneura + Caenogastropoda). An overall compatible topology is found by bootstrap using parsimony of the same data (DNABOOT, figure 3), in spite of minor differences within Euthyneura and of the seemingly better resolution within the Caenogastropoda, which however is not well supported by bootstrap proportions. After modification of the composition of the sample, reduced to twenty species and keeping only *Acanthochitona* as an outgroup, a bootstrap based on the same data still provides a similar topology (DNABOOT, figure 4). Finally the replacement of *Acanthochitona* by *Mytilus* as the outgroup in the latter data set, provokes the shift of *Patella* to the sister-taxon of the Vetigastropoda (i.e. monophyly of the Archaeogastropoda, figure 5). In addition to these changes, the shift of *Rhynchopelta* as the sister-taxon of all other gastropod taxa in the data set may be found by using NJBOOT instead of DNABOOT (not shown).

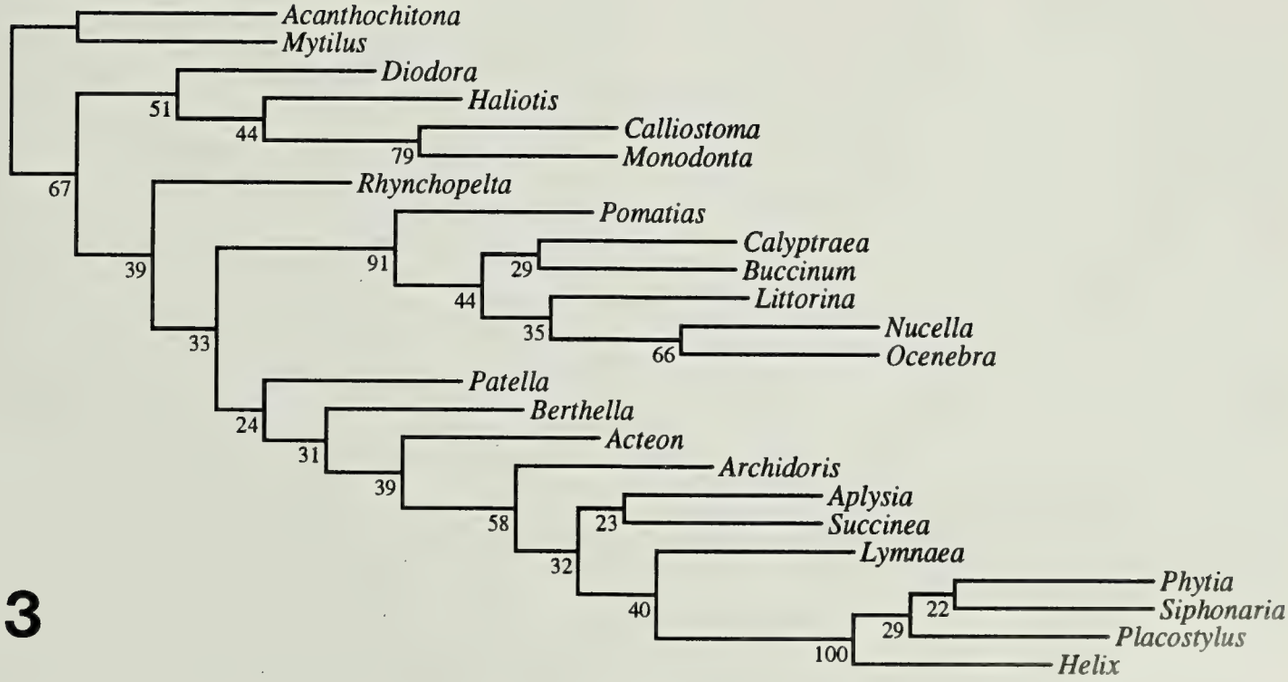
Shifts in the position of several taxa, similar to those induced by changes in the taxonomic composition of the sample, may easily be provoked by minor changes in the alignment of the sequences, especially when variable regions are included (not shown). It is precisely those taxa whose relationships are uncertain which shift the most easily.

However, nodes that are modified by such minor changes in the data set have bootstrap proportions that are always much lower than 90%, whereas nodes with BP that are the closer to 90% remain unchanged (mono-

Table 4. Relative rate test using either *Acanthochitona* or *Mytilus* as an outgroup. Distances in % differences (sites informative for parsimony), variable regions overlined in table 1 excluded.

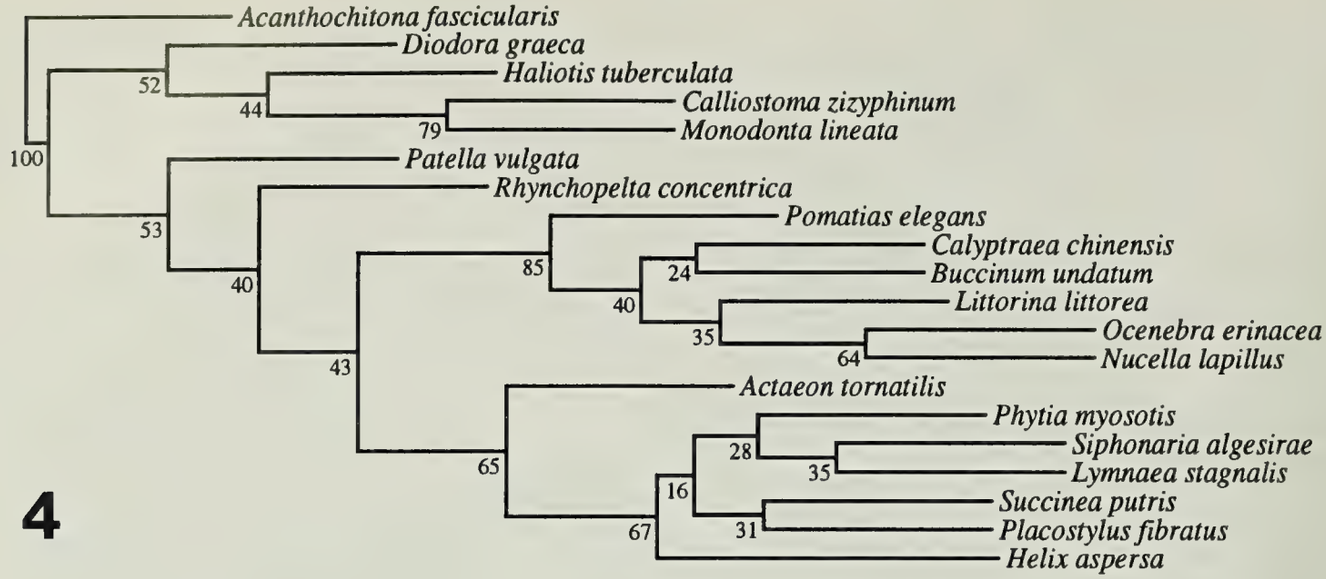
	Helix	Placo	Phyti	Sipho	Lymna	Succi	Aplys	Archi	Actae	Berth	Ocene
Acanthochitona	64.52	65.00	68.85	74.19	65.57	63.93	67.74	61.29	58.06	59.65	59.02
Mytilus	60.32	57.38	62.30	66.13	59.68	58.06	58.73	57.14	55.56	61.40	59.68
	Nucel	Bucci	Calyp	Pomat	Litto	Rhync	Monod	Calli	Halio	Diodo	Patel
Acanthochitona	59.68	62.30	64.52	60.00	65.57	51.61	54.84	55.56	55.00	54.84	63.49
Mytilus	58.73	60.66	61.90	59.02	60.66	50.79	49.21	54.69	65.00	56.25	73.44

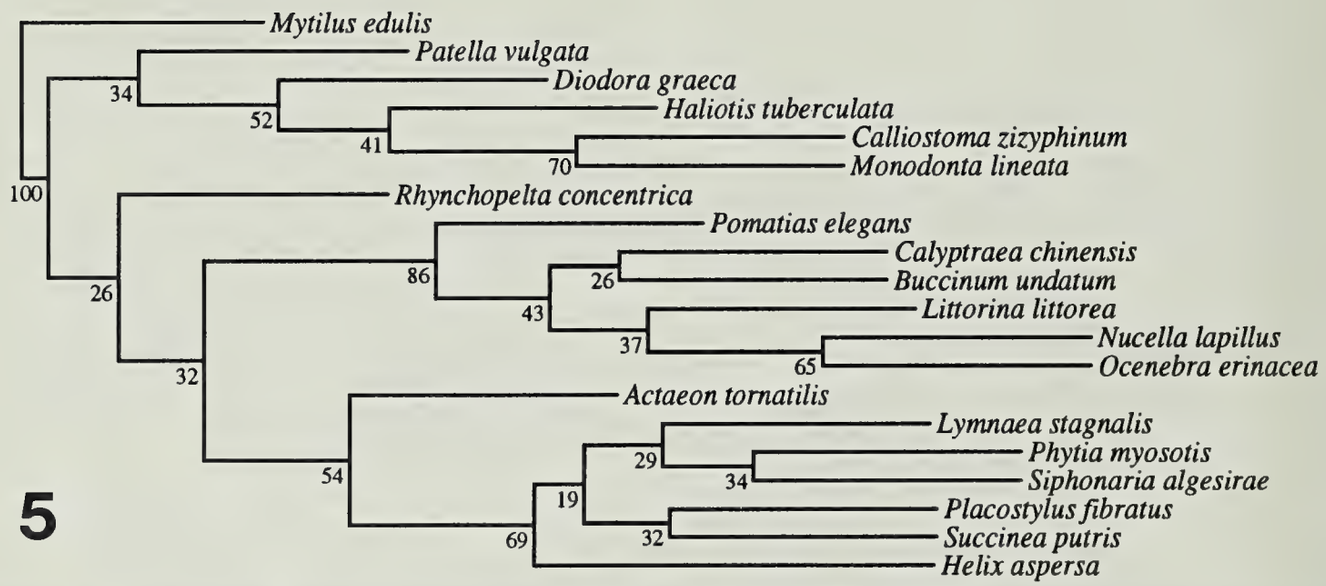
2



3

Figures 2–3. Parsimony tree (figure 2) and bootstrap proportions based on parsimony trees (figure 3) of the whole data set (table 1, without *Mus*). **2.** Consensus of the 13 shortest trees obtained with Hennig86, options m*; bb*; nelsen. Data table 1, overlined regions not taken into account. **3.** Topology and bootstrap proportions of majority nodes obtained by parsimony (DNABOOT) from the same data. BP in %, branch lengths in % of 1000 replicates (terminal branches = 100).





Figures 4–5. Topology and bootstrap proportions in % (BP), calculated with DNABOOT from incompete samples. **4.** Trees obtained for 20 taxa , from the same alignment and regions as in figures 2 and 3, *Acanthochitona* as the outgroup. **5.** Tree obtained from the same data set, but with *Mytilus* substituted for *Acanthochitona* as the outgroup. BP in %, branch lengths in % of 1000 replicates (terminal branches = 100).

phyly of Vetigastropoda and subgroups; monophyly of Caenogastropoda). This observation confirms previous empirical recommendations (e.g. Felsenstein, 1985), and leads us to regard as unreliable most of the nodes of the trees displayed in figures 2, 3, 4 and 5. The obvious solution to this problem is to use more taxa and longer sequences, which would allow the use of more homologous sites. However, for both practical (cost!) and theoretical reasons discussed in the previous section, we consider it more useful to analyze further within our data set which information may, or may not, be considered reliable.

Monophyly of the Gastropoda and Relationships of the Patellogastropoda

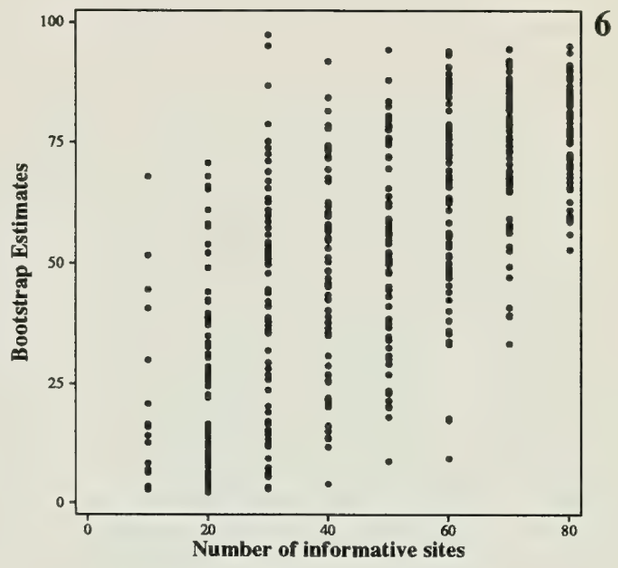
Using the complete data set, jackknives on increasing number of sites correlate with bootstrap values increasing

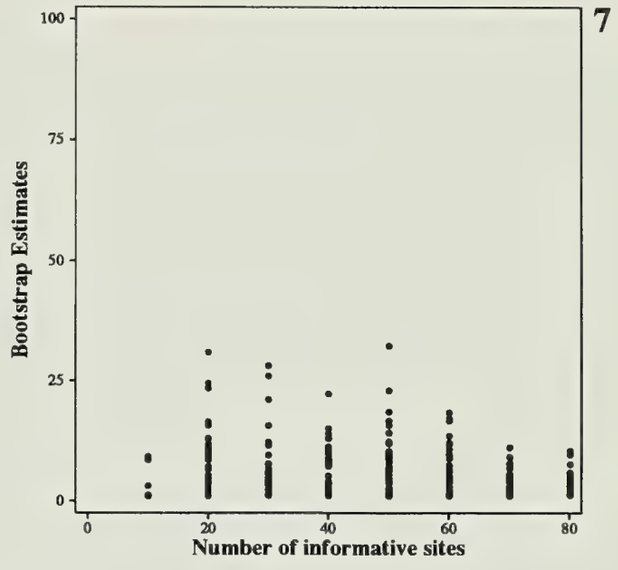
in mean and decreasing in amplitude of variation (figure 6). The 84% BP obtained with 93 sites will probably increase when more sites are used, and the monophyly of the Gastropoda may be considered as established, at least when Polyplacophora and Bivalvia are used as outgroups.

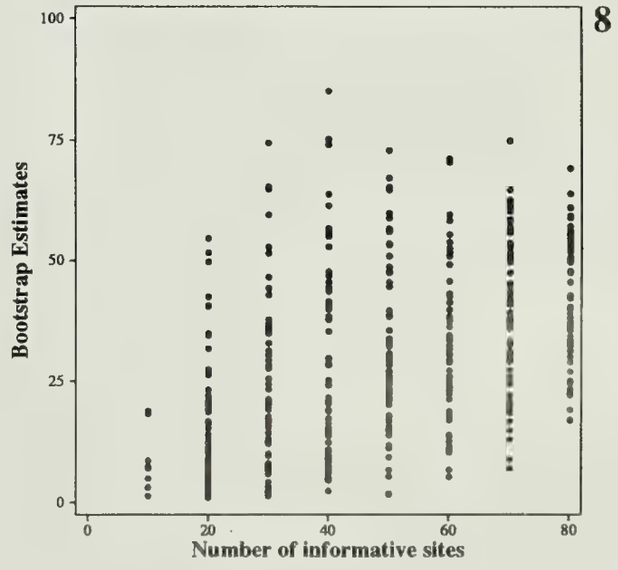
The relationship of *Patella* to other gastropod taxa is more difficult to establish. Three possibilities exist: (a) *Patella* is the sister-group of other Gastropod taxa, a position defended by Haszprunar (1988) and Lindberg (1988); (b) *Patella* and the Vetigastropoda form a monophyletic taxon, the "Archeogastropoda" in their classical definition and position; (c) *Patella* and the Caenogastropoda + Euthyneura form a monophyletic group. In our data set, the bootstrap proportions of the grouping (c) seem to increase with the number of sites sampled; however, no value higher than 53% has been found, and the mean bootstrap value for 80 sites is less than 20%; furthermore this grouping has been found only 243 times (of 1152) with bootstrap proportions not exceeding 15% in our "jackmono" results. Consequently we reject this grouping.

More surprisingly, the increase and variation of BP in relation with number of sites sampled seems to favour monophyly of *Patella* with the Vetigastropoda: BP of the group *Patella* + *Acanthochitona* + *Mytilus* tend to decrease in both mean and amplitude of variation as the number of sites is increased (figure 7), and never exceed 32%; whereas BP of the group *Patella* + Vetigastropoda seem to increase in mean and decrease in variation as the number of sites sampled is increased, and reach 85% (figure 8). However, the solution is not simple, as the mean value of BP for 80 sites sampled is only ca. 40%, which is not significant.

The result of BP calculated from combinations of seven species from the seven higher taxa defined above is even more ambiguous, and somewhat contradictory (figures 9 and 10). For both groupings, the histograms are bimodal, the bimodality being far more obvious in figure 10 (Archeogastropoda monophyletic) than in figure 9 (*Patella* = sister-group of other gastropods). The monophyly of the Archeogastropoda is supported 531 times of 1152, and reaches BP as high as 952, i.e. a value worth being considered; the paraphyly of the Archeogastropoda is

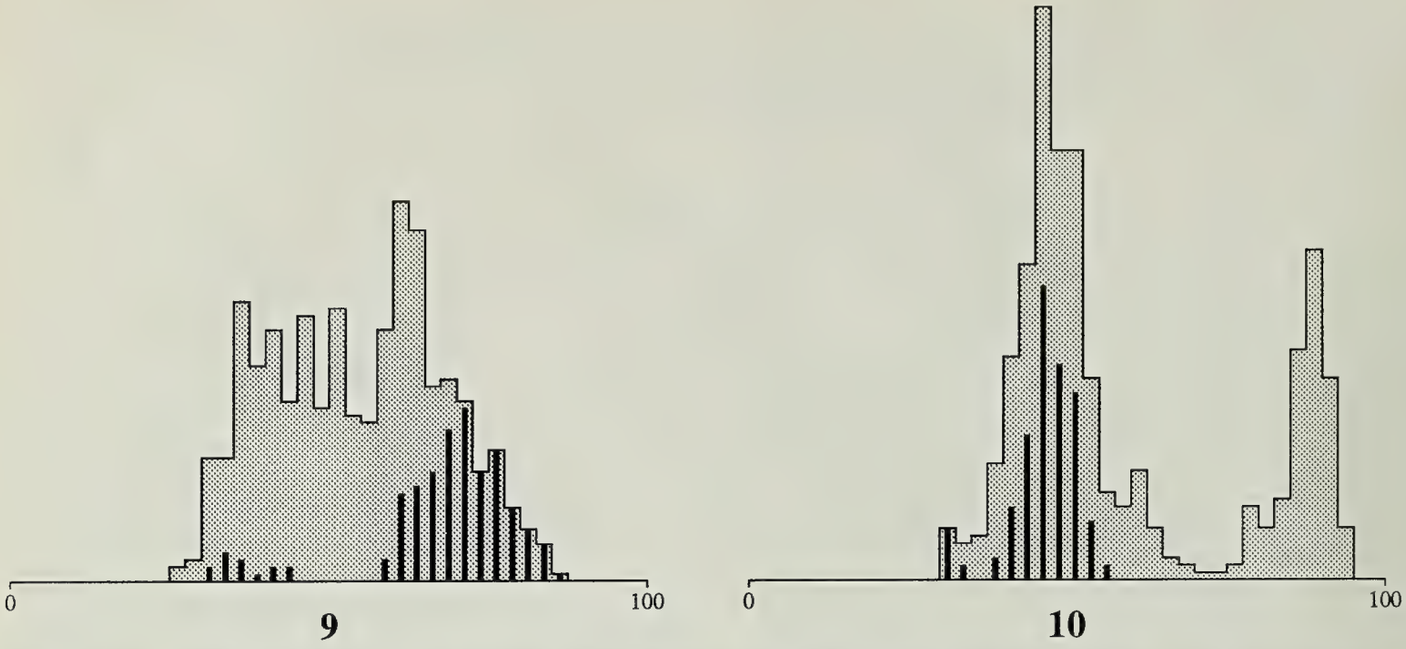






→

Figures 6–8. BP in relation to number of sites jackknived, data set from table 1, *Mus* excluded. **6.** Grouping of all Gastropoda. **7.** Grouping of all Gastropoda except *Patella*, i.e. for the position of the Patellogastropoda (*Patella*) as the sister-group of the other Gastropoda. Note that BP do not increase as site number is increased. **8.** BP in relation to number of sites jackknived for the grouping of *Patella* with the Vetigastropoda (*Diodora*, *Haliotis*, *Calliostoma*, *Monodonta*), i.e. for the monophyly of the Archaeogastropoda: note that BP increase as the number of sites sampled is increased. Number of sites jackknived along the abscissa, BP in % along the ordinate; number of jackknife replicates 70 for each number of sites, number of bootstrap replicates of each jackknife replicate 1000.



Figures 9–10. Histograms of BP of monospecific combinations of higher taxa for two positions of *Patella*, from data set table 1, overlined regions not taken into account. **9.** Grouping of the outgroup with *Patella*, i.e. the position of the Patellogastropoda as the sister-group of the other Gastropoda (found 585/1152 times); BP for combinations including *Mytilus* shown in dark. **10.** grouping of *Patella* with the Vetigastropoda, i.e. the monophyly of the Archeogastropoda (found 531/1152 times); BP for combinations including *Acanthochitona* shown in dark.

supported 585 times, with BP not exceeding 852, i.e. a value too low to be considered highly significant, but which could be raised by additional data. These unclear results may be the consequence of long branch attract: both *Patella* and *Diodora* exhibit rates of substitution higher than in the other species of the sample.

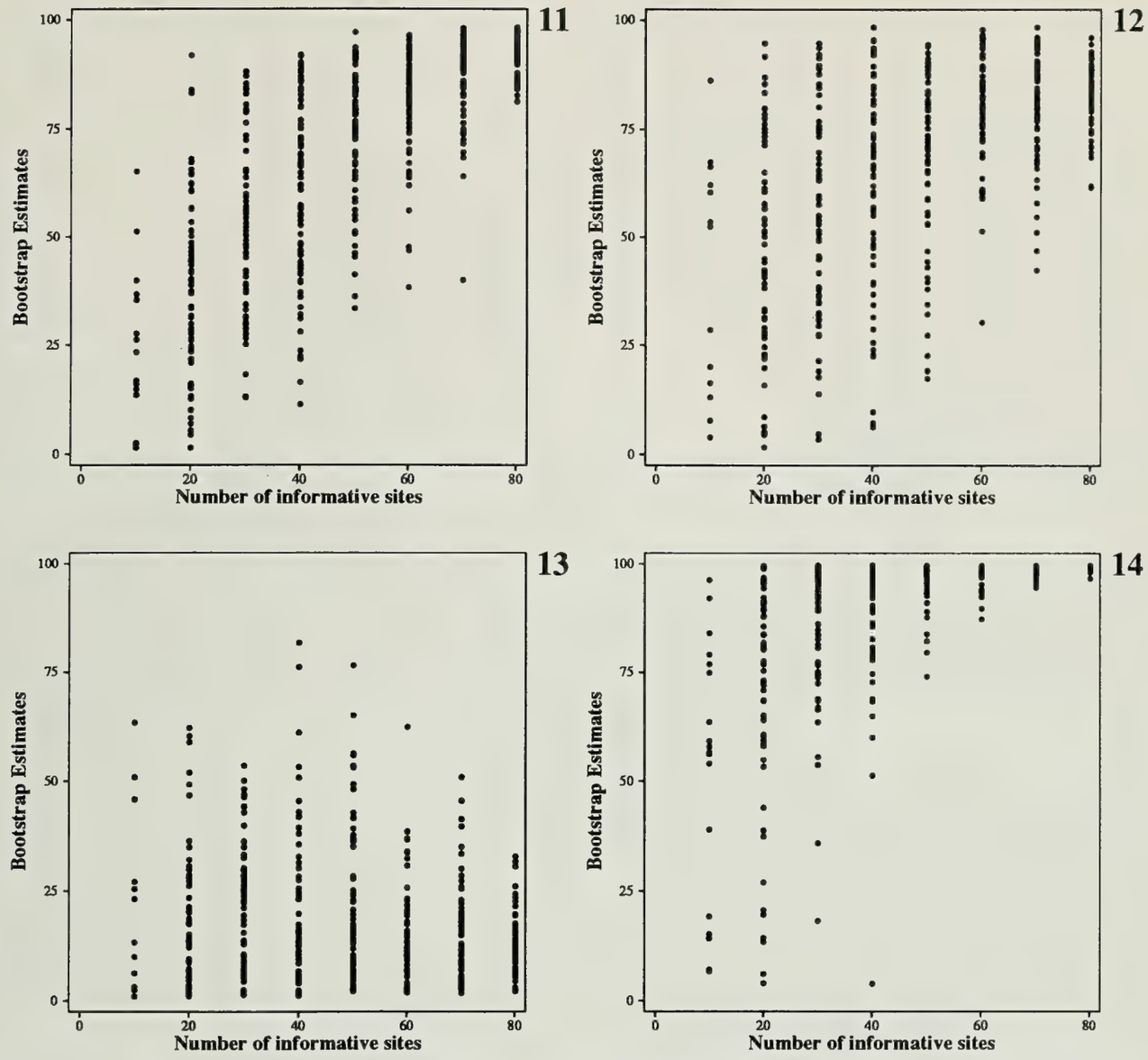
The analysis of the distribution of taxa in the histograms shows that high BP for monophyletic Archeogastropoda are equally supported by all of the 24 species except *Acanthochitona*, *Diodora* and *Calliostoma* which support low BP exclusively: in other words, one may obtain apparent monophyly of the Archeogastropoda at a significant level of confidence by using another outgroup than *Acanthochitona* and by removing *Diodora* and *Calliostoma* from the sample. The paraphyly of the Archeogastropoda is favoured by using *Mytilus* as an outgroup, is contradicted by *Diodora*, but is strongly supported by *Monodonta*, whereas *Acanthochitona* does not influence this node. Several explanations for the strong bimodality of the Archeogastropod node may be found: although it is improbable when considering BP discussed further for Vetigastropoda, *Diodora*, which contradicts both solutions, could belong to a clade other than the Vetigastropoda; or the relative rates of substitution may be involved: *Patella* and *Diodora* may aggregate because both evolved faster (long branches attract), while the influence of *Monodonta* may be explained by its slower relative rate when *Mytilus* is taken as an outgroup (table 4).

It would be premature to reach any definitive conclusion about the relationship of the Patellogastropoda, which may be considered either as: (a) members of a monophyletic Archeogastropoda, a solution that the present results favour slightly, but not enough to reject the hypothesis that it is due to long branches attract; or (b) as the sister-group of the other gastropods or even a less related taxon, a solution that is strongly favoured by the morphological analyses cited above.

Monophyly and Relationships within the Vetigastropoda

As discussed by Salvini-Plawén & Haszprunar (1987) and Haszprunar (1988), the Vetigastropoda are characterized morphologically by synapomorphic sensory "bursicles" located at the efferent edge of each ctenidial leaflet; common and distinctive types of tentacles and anterior oesophagus epithelium; epipodial sense organs; high chromosome number; and similar renal cell-types. Within the Vetigastropoda, the Pleurotomarioidea and Trochoidea, represented in this study by *Haliotis* and *(Calliostoma* + *Monodonta)* respectively , are united by aberrant osphradial characters. This implies that the Fissurelloidea, here represented by *Diodora*, are the sister-group of (Pleurotomarioidea + Trochoidea).

Our data fully confirm this phylogenetic arrangement, and contradict the more classical view, in which the Pleurotomarioidea (*Haliotis*) are the sister-group of the Fissurelloidea + Trochoidea. The latter view was defended for example by Hickman & McLean (1990) and supported by fewer sequences and other methods (Tillier et al., 1992). Indeed BP on jackknives of various numbers of sites strongly support, with little doubt: (a) the monophyly of the Vetigastropoda (figure 11); (b) the mono-

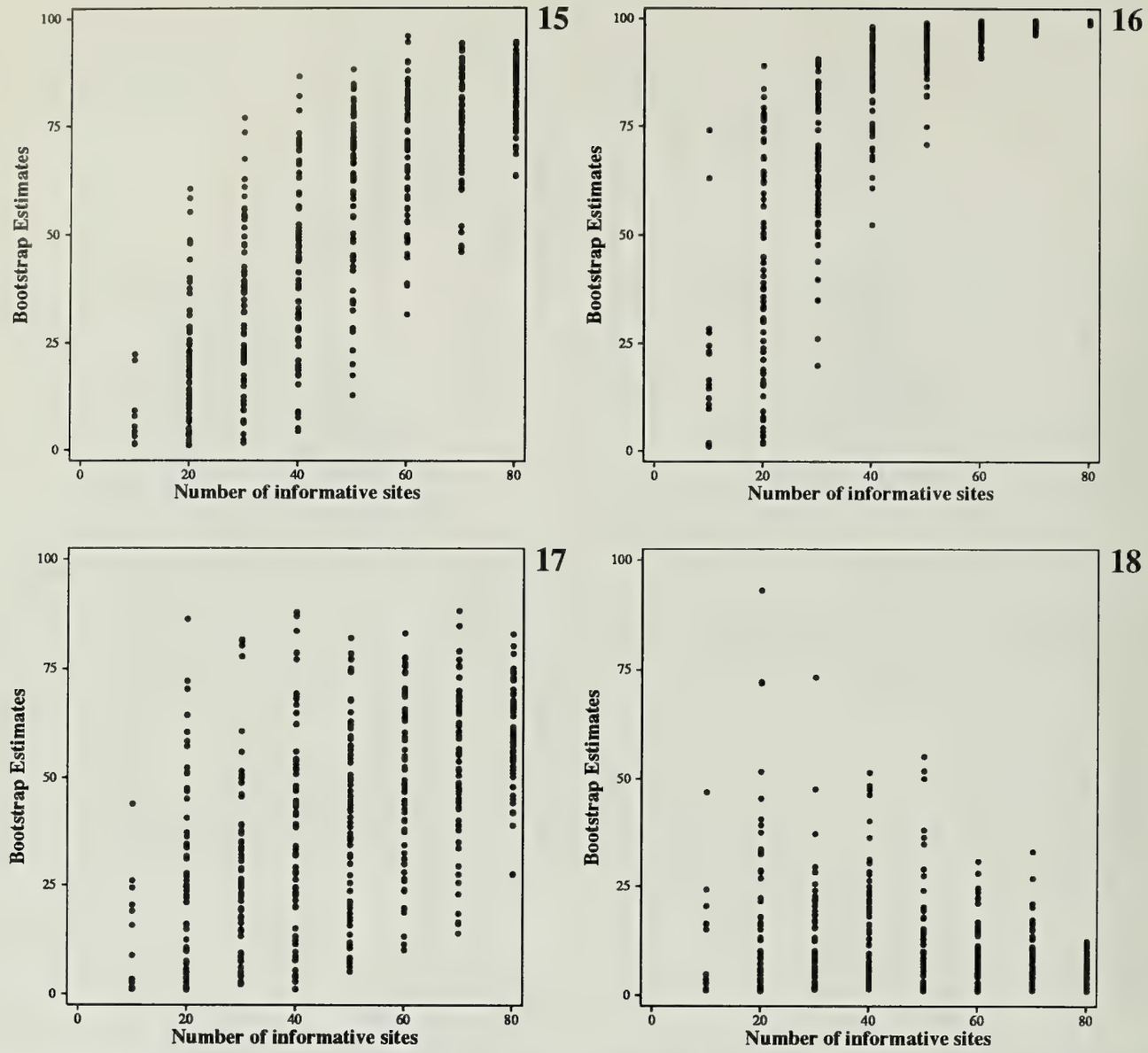


Figures 11–14. BP in relation to number of sites jackknived from the data set of table 1. **11.** Grouping of the Vetigastropoda together, i.e. monophyly of the latter (*Diodora*, *Haliotis*, *Calliostoma*, *Monodonta*). Note that BP increase as site number is increased. **12.** Grouping of the Pleurotomarioidea (*Haliotis*) with the Trochoidea (*Calliostoma* and *Monodonta*), i.e. Fissurelloidea as the sister group of the latter. Note that BP increase as site number is increased. **13.** Grouping of the Fissurelloidea (*Diodora*) with the Trochoidea (*Calliostoma*, *Diodora*), i.e. Pleurotomarioidea (*Haliotis*) as the sister group of the latter. Note that BP do not increase, but rather diminish as site number is increased. **14.** Grouping of the Trochoidea, i.e. monophyly of the latter. Note that BP may reach 100%, and vary very little, for 80 sites. Number of sites jackknived along the abscissa , BP in % along the ordinate; number of jackknife replicates 70 for each number of sites, number of bootstrap replicates of each jackknife replicate 1000.

phyly of *Haliotis* and Trochoidea (figure 12) and refute the monophyly of *Diodora* and the Trochoidea (figure 13); (c) the monophyly of the Trochoidea *Calliostoma* and *Monodonta*, depicted in figure 14 for the sake of showing the pattern produced by a nearly fully resolved node.

Monophyly of Apogastropoda, Caenogastropoda, Stenoglossa

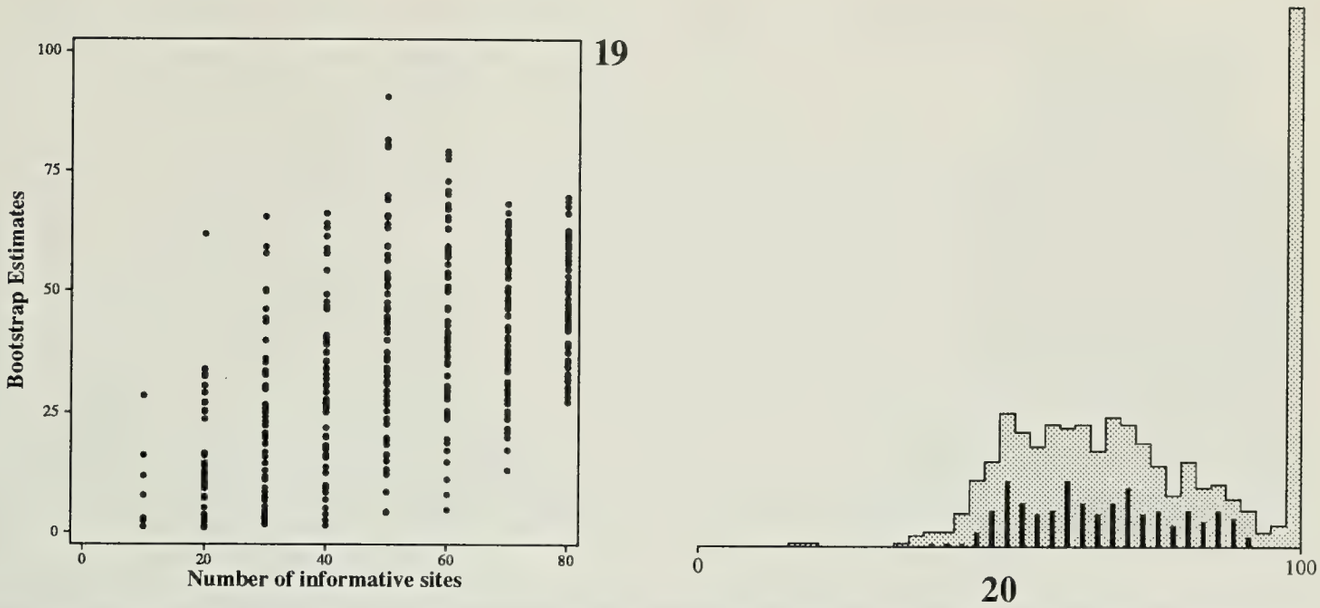
Surprisingly, the monophyly of the Apogastropoda, i.e. Caenogastropoda plus Euthyneura in our sample (figure 1), is not among the best supported groupings, although an increasing BP with increasing number of sites sampled indicate that this node could be confirmed by ca. 40 more informative sites (figure 15). The histograms of BP for monospecific samples of each higher taxon are unimodal (figure 23) and reach higher values when hypervariable regions are retained (maximum 968 vs. 926, mean 764 vs. 386). Analyses of the corresponding histograms for BP without hypervariable regions show that BP are lowered by *Acteon* (Cephalaspidea) and *Berthella* (Notaspidea). As noted previously (Tillier et al., 1992) and visible in table 1, opisthobranch D1-C1 sequences



Figures 15–18. BP in relation to number of sites jackknived from the data set of table 1. **15.** Grouping of the Apogastropoda (Caenogastropoda + Euthyneura, figure 1), *Rhynchopelta* excluded. BP increase as site number is increased. **16.** Grouping of the Caenogastropoda (cf. figure 1). BP increase as site number is increased, and the maximum BP (100%) is reached and exhibits little variation for 80 sites. **17.** Grouping of *Calyptraea* and Stenoglossa (*Buccinum*, *Ocenebra*, *Nucella*). BP increase as site number is increased, but the mean BP reached for 80 sites is far from highly significant values. **18.** Grouping of the Stenoglossa (*Buccinum*, *Ocenebra*, *Nucella*). BP decreases as site number is increased, suggesting possible polyphyly of the latter; compare with figure 17. Number of sites jackknived along the abscissa , BP in % along the ordinate; number of jackknife replicates 70 per number of sites, number of bootstrap replicates of each jackknife replicate 1000.

tend to differ more than usual from other gastropod sequences, inducing unexpected fluctuations in BP. However, figure 15 suggests that longer sequences could resolve this problem at least at the level of the (Caenogastropoda + Euthyneura) node. The polyphyly of the Euthyneura cannot be excluded (see further).

The monophyly of the Caenogastropoda is very well supported by our data (figure 16), although supported thus far by a single synapomorphic morphological character, a highly distinctive type of osphradium (Haszprunar, 1988a). The inclusion of the Stenoglossa within Caenogastropoda is strongly supported by our sample, and we reject the hypothesis of a non-caenogastropod origin of the group (Ponder, 1973; Taylor & Morris, 1988). The monophyly of the Muricidae *(Ocenebra* and *Nucella*), as well as the monophyly of the Littorinoidea *(Littorina* and *Pomatias*), are well supported (not shown). More interesting is the lack of support for the monophyly of Stenoglossa (*Buccinum* + *Ocenebra* + *Nucella*) (figures 1 to 4), which may not be an artifact: surprisingly, the inclusion of *Calyptraea* with the Stenoglossa is better supported (figure 17) than the monophyly of the Stenoglossa without *Calyptraea* (figure 18), which could indicate the polyphyly of the Stenoglossa. The monophyly



Figures 19–20. BP in relation to number of sites jackknived, and histogram of BP of monospecific combinations of higher taxa, for the grouping of the Euthyneura (data set table 1, currently admitted relationships figure 1), overlined regions not taken into account in figure 20). **19.** Dot diagram: although BP increase as site number is increased, mean value for 80 sites is not highly significant at all. Number of sites jackknived along the abscissa, BP in % along the ordinate; number of jackknife replicates 70 per number of sites, number of bootstrap replicates of each jackknife replicate 1000. **20.** Histogram from JACKMONO: BP for combinations including *Acteon* shown in dark. *Archidoris* and *Berthella* correspond to the same distribution as *Acteon*; on the contrary *Aplysia* occurs only in combinations forming the peak on the extreme right.

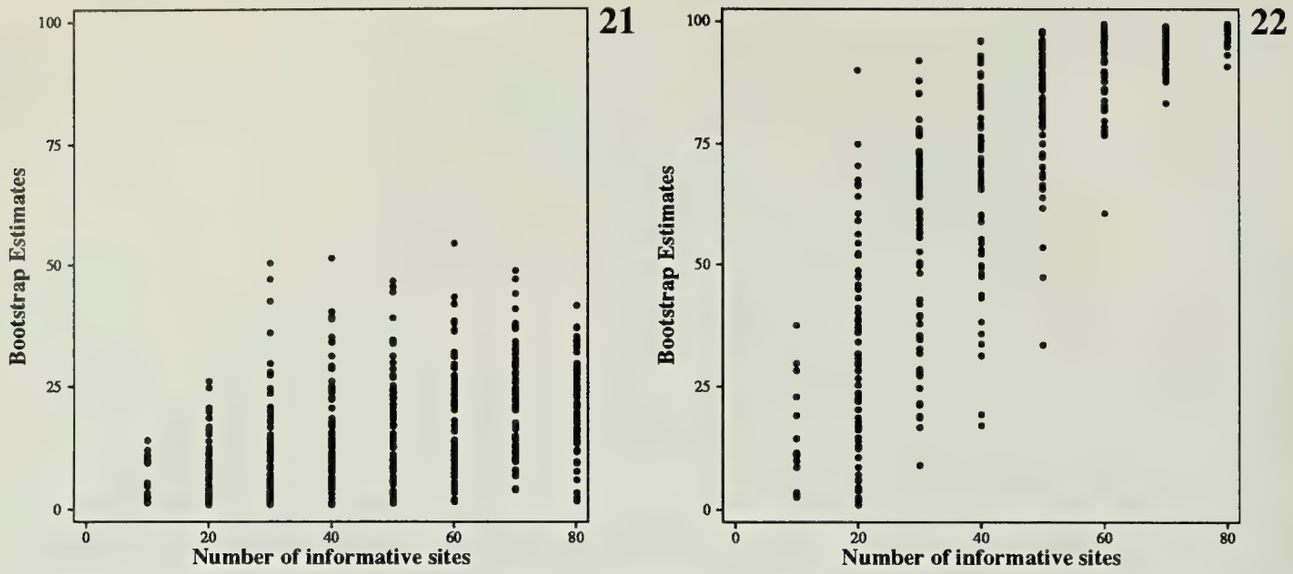
of *Calyptraea* and *Buccinum* cannot be rejected without additional data (its BP increases slowly in relation to number of sites to reach a mean of 380 for 80 sites; maximum value 918, which is not neglegible). The paucity of taxa and number of sites in our sample precludes any further comment other than underscore the need to enlarge the data set for Caenogastropoda and Stenoglossa.

Monophyly of, and relationships within the Euthyneura

As with the Caenogastropoda and for the same reason, the monophyly of Euthyneura is not as well supported as expected, considering that their monophyly based mainly upon their pentaganglionate condition has not been contested. Their BP increase with number of sites sampled (figure 19), but are still far from significant values for 80 sites. The analysis of the strongly bimodal histogram of BP for monospecific samples of each higher taxon shows that (figure 20): (a) the Euthyneura are very strongly supported when *Aplysia* represents the Opisthobranchia (BP between 95 and 100%); (b) other Opisthobranchia occur only in combinations corresponding to the lower mode, with BP between ca. 400 and ca. 900. As for Apogastropoda, our sequences of non-Anaspidea Opisthobranchia do not support monophyly of the Euthyneura, either because of their autapomorphies or because the Opisthobranchia are polyphyletic, but not because they are too recent (at least one more external node is well supported, see next paragraph). Their relative rates of substitution do not appear sufficiently different from the others to explain alone this lack of resolution (tables 3, 4). Until more and longer sequences allow the choice of either answer, several points may be made.

Within the Euthyneura, our data do not support clearly the monophyly of Pulmonata alone in the total sample (figure 21), but support even less their para- or polyphyly (not shown). Our data do support strongly the monophyly of *Aplysia* + Pulmonata (figure 22). Removing *Aplysia* from the sample immediately raises the maximum observed BP for Pulmonata from 544 to values between 900 and 1000. However, it cannot be concluded that the Pulmonata are para- or polyphyletic since, in our data set, no grouping within the Anaspidea + Pulmonata clade shows a clear trend toward stability similar to that of the clade as a whole. The hypothesis of the monophyly of the Pulmonata remains the most probable considering morphological characters (e.g. Tillier, 1984), but combination of morphological and molecular data presented suggests the position of the Anaspidea as the sister-group of the Pulmonata (with the proviso that only four of the nine opisthobranch orders recognized, for example, by Boss (1982), are included in our sample). To pursue further the analysis of this clade will require additional molecular characters as well as re-evaluation of the morphological data.

Our data do not provide any clear further indication on relationships of the Cephalaspidea (*Acteon*), Notaspidea (*Berthella*) and Nudibranchia (*Archidoris*). The groupings that seem the most promising are Notaspidea + Nudibranchia (BP reaching 932, but mean for 80 sites



Figures 21–22. BP in relation to number of sites jackknived for the grouping of the Pulmonata with (22) and without (21) *Aplysia* (currently admitted relationships figure 1). **21.** Without *Aplysia*, BP do not increase significantly as site number is increased. **22.** *Aplysia* included, BP increase as site number is increased, and reach highly significant values for 80 sites. Number of sites jackknived along the abscissa , BP in % along the ordinate; number of jackknife replicates 70 per number of sites, number of bootstrap replicates of each jackknife replicate 1000.

ca. 400!) and Cephalaspidea + Anaspidea + Pulmonata (BP reaching 769, but mean for 80 sites ca. 200 only). Neither monophyly of the Opisthobranchia less *Aplysia*, nor paraphyly or even polyphyly of this taxon may be rejected. At the moment, a paraphyletic arrangement between emergence of the Caenogastropoda and that of the Anaspidea + Pulmonata clade appears to be the most reasonable when trying to synthetize available information, in agreement with Haszprunar's (1988c) framework. However, the hypothesis of the early emergence of the Euthyneura proposed by Ponder & Warén (1988: 291) is worth investigating further.

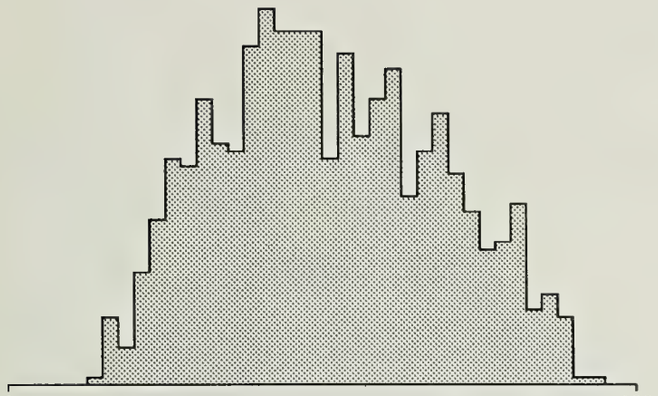
Relationships of the Neomphalina

As could be expected from previous morphological analyses (see Haszprunar, 1988a, b, c; McLean, 1990b), *Rhynchopelta* always roots very deeply within the trees obtained from our data set. In addition to the fact that it is the only representative of its group (due to the difficulty in obtaining RNA from deep-sea gastropods), its position makes its relationships still more difficult to analyze because of the influence of proximal nodes (mentioned above and illustrated by the shift provoked by the change of the outgroup (figures 4 & 5)). The inclusion of *Rhynchopelta* in the group (outgroups + Patellogastropoda + Vetigastropoda) is supported by the monophyly of the Apogastropoda (figure 15), and is not contradicted by the histogram of monospecific combinations of higher taxa (figure 23). However, these data do not distinguish between the four possible solutions (two of them shown by figures 1 and 4, plus monophyly of *Rhynchopelta* + Vetigastropoda, or monophyly of *Rhynchopelta* + Euthyneura).

As in the case of the Patellogastropoda, the variation of BP as a function of an increasing number of jackknived sites is not conclusive, but suggests rejection of the presently favoured solution: the grouping (outgroup + *Patella* + *Rhynchopelta*). This grouping, with Neomphalina emerging between the outgroup and the Vetigastropoda, as advocated by Haszprunar (1988a, b, c), has not been found to have a BP higher than 1%. The monophyly of *Rhynchopelta* and *Patella* is sometimes found, but too rarely (293/1152) and with too low BPs (mean ca. 1.5% for 80 sites!) to be retained, as confirmed by its absence from BP of monospecific combinations of

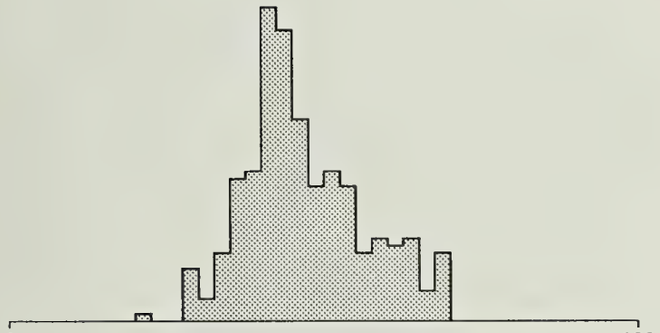
→

Figures 23–28. Histograms of BP of monospecific combinations of higher taxa, with (23, 25, 27) and without (24, 26, 28) *Patella*. **23–24.** Grouping of the Apogastropoda (Caenogastropoda + Euthyneura, figure 1), excluding *Rhynchopelta*. The high BP of figure 23 are reinforced when *Patella* is removed from the sample (figure 24). **25–26.** Grouping of *Rhynchopelta* with the outgroup (including *Patella* in figure 25), with samples including (figure 25) or not including (figure 26) *Patella*. These combinations have been found only 277 and 152 (of 1152) times. **27–28.** Grouping of *Rhynchopelta* with the Apogastropoda, with samples including (figure 27) or not including (figure 28) *Patella*. These combinations have been found 541 and 979 (of 1152) times; comparison with figures 27 and 28 shows the much better support for monophyly of Neomphalina, Caenogastropoda and Euthyneura (compare with figure 1).

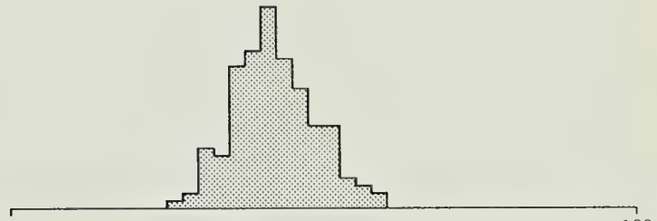


23

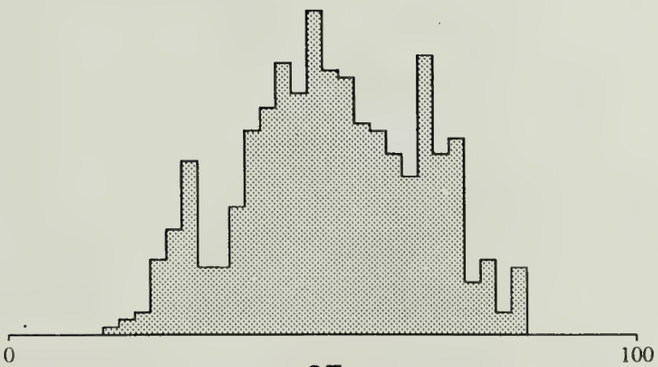
24



25



26



27

28

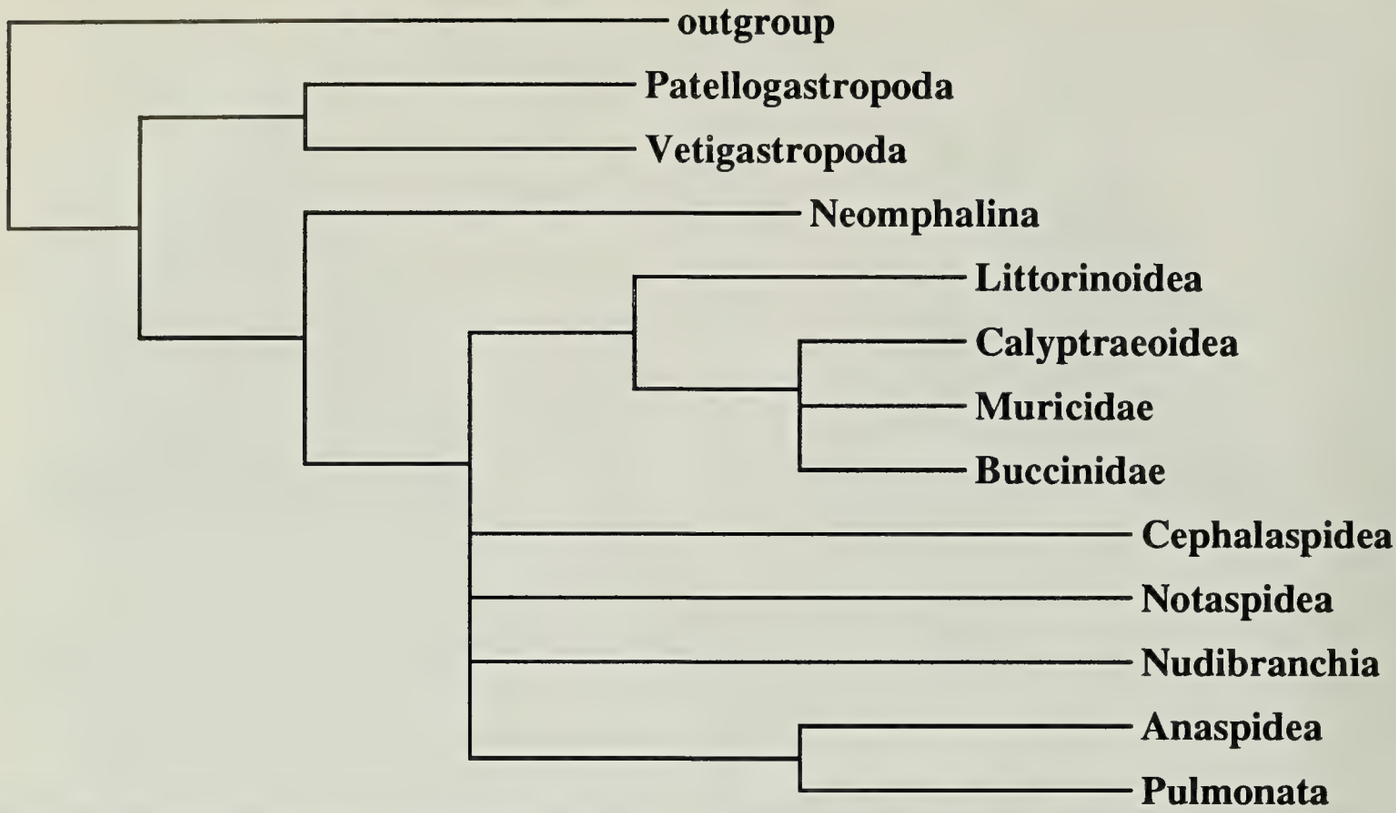


Figure 29. Diagram showing the best supported relationships from data set of table 1; compare with figure 1. Monophyly of Patellogastropoda + Vetigastropoda (= Archaeogastropoda) is found better supported, although not conclusively, than their paraphyly (long branch attraction); monophyly of the Opisthobranchia (Cephalaspidea + Notaspidea + Nudibranchia + Anaspidea) is not supported at all, whereas monophyly of Anaspidea and Pulmonata is supported. Neomphalina (*Rhynchopelta*) are found the sister-group of the Caenogastropoda + Euthyneura included in the sample.

higher taxa. We are therefore left with three possible solutions: (a) *Rhynchopelta* belongs to supposedly monophyletic Archeogastropoda, either as the sister-group of (*Patella* + Vetigastropoda), as in figure 1, or as the sister-group of the Vetigastropoda; (b) *Rhynchopelta* and the Vetigastropoda form a monophyletic unit but the Archeogastropoda are paraphyletic; (c) *Rhynchopelta* and the Apogastropoda (Caenogastropoda + Euthyneura) are sister-groups, as in figure 4. None of these solutions has produced a mean BP greater than 25% for eighty sites. To analyze further our limited data set, it is more reasonable to turn to monospecific combinations of higher taxa, which equilibrate the respective samplings of the latter, to try to find the most reliable solution.

Bootstraps have been calculated in monospecific combinations of higher taxa both with and without *Patella*; allowing us (a) to consider only one problematic monospecific higher taxon at a time; (b) to eliminate one problematic taxon that not only is close to the taxon under question and therefore influences the node(s) analyzed, but also has a higher substitution rate than average, while *Rhynchopelta* has a lower than average rate (see figures 23 and 24 for comparison). Comparisons of the BP obtained for the various possible nodes (figures 25 to 28) indicate that in our sample:

(a) *Rhynchopelta*, and thus presumably the Neomphalina, is probably not an outgroup to the Vetigastropoda + Apogastropoda clade (figures 25–26);

(b) *Rhynchopelta*, *Patella* and the Vetigastropoda do not constitute a monophyletic unit (this grouping never found with a BP higher than 40%);

(c) *Rhynchopelta* and the Vetigastropoda do not constitute a monophyletic unit (this grouping never found with *Patella* included in the sample, found 7/1152 times with a maximum BP of 43% with *Patella* excluded from the sample);

(d) the monophyly of *Rhynchopelta* and the Apogastropoda is the best supported grouping, as shown by figures 27 (*Patella* included) and 28 (*Patella* not included).

Whether Peltospiridae and Neomphalidae are family-rank taxa in a superfamily Neomphaloidea as proposed by Warén & Bouchet (1989), or are representatives of two superfamilies in an order Neomphalina, as maintained by McLean (1990b), their close relationships are currently admitted despite the lack of known synapomorphies. They have been reported to originate: (a) in all cases, outside of the clade Vetigastropoda although being "at an archeogastropod level of organization" (Fretter, 1989); (b) between the outgroup and the Vetigastropoda (Haszprunar, 1988a, b, c); or (c) among the

Architaenioglossa ("Viviparifórmes" Sitnikova & Starobogatov, 1982; Golikov & Starobogatov, 1988). In relation to Haszprunar's (1988c) phylogenetic framework (figures 1, 29), support here found for the emergence of the Peltospiridae (and Neomphalidae?) between the Vetigastropoda and the Caenogastropoda, confirm the first hypothesis, refutes the second, and is compatible with the third—which has been by far the least advocated. However, no more definitive conclusion can be made before data similar to those analyzed here are available for both more Neomphalina and some Architaenioglossa.

CONCLUSION

In spite of this relative low level of resolution, the methods elaborated by Lecointre et al. (1993 and in press) prove more powerful than "simple" jackknives and bootstraps for such limited data sets as ours. Under such conditions the power of these methods is doubtlessly increased when they are used in a comparative way to choose one among a finite number of solutions, as done here, rather than to estimate the absolute reliability of any single solution for a given data set—even if this is what they were designed for, and how they should be used as often as possible. Choosing the best among unequally poor solutions, as often done here, is a gamble—it is not equivalent to finding the exact solution. However, this estimation is justified in that, overall, our results converge with those obtained from morphological analyses. The methods used here may help to answer the molecular systematist's dilemma: more taxa, or more characters?

Combining the nodes retained above, i.e. the phylogenetic hypotheses the best supported by our data set, results in the tree shown in figure 29: (a) the Anaspidea and Pulmonata form a monophyletic unit; (b) the position and relationships of the other Opisthobranchia orders are doubtful; (c) the Caenogastropoda are monophyletic, but no reliable support has been found for monophyly of the Stenoglossa families Buccinidae and Muricidae—which however was not rejected either; (d) the Neomphalina are the sister-group of the Apogastropoda; (e) the Vetigastropoda are monophyletic, while the Fissurelloidea are the sister-group of the Pleurotomarioidea + Trochoidea; (f) the Patellogastropoda and Vetigastropoda may form a monophyletic unit Archeogastropoda, which however must be tested by addition of taxa.

All these hypotheses are not equally well supported: the position of the Patellogastropoda is dubious and must be confirmed by additional data from at least one or two more taxa from this same group, as well as possibly by additional sequence length. Given the shape of the BP/number of sites diagrams, additional sequences are required to increase the reliability of most of the other proposed relationships: in nearly all cases discussed but the Trochoidea and Caenogastropoda, more characters could yield highly significant results in a non-comparative way. Data of the same level of variability would help to resolve the Euthyneura problem, whereas sequences of higher variability would possibly solve the Pulmonata and Stenoglossa questions.

ACKNOWLEDGMENTS

We are greatly indebted to Hervé Philippe for providing the programs used and to Guillaume Lecointre, Hervé Le Guyader and André Adoutte for discussion, access to their unpublished manuscripts and comments on the present paper. Bootstraps using DNABOOT have been done in the Centre Informatique du Muséum, MNHN. We owe to Serge Gofas, Carmen Salas and to the staff of the Roscoff marine station part of the material used. Anders Warén sorted and determined the hot vent limpets kindly provided by Michel Segonzac (IFREMER, cruise HERO 91, chief Daniel Desbruyères).

LITERATURE CITED

Barr, P. J., R.M. Thayer, P. Laybourn, R.C. Najarian, F. Seela and D.R. Tolan. 1986. Deaza -2'- deoxyguanosine -5'-triphosphate: enhanced resolution in M13 dideoxy sequencing. Biotechniques 4: 428–432.

Boss, K. J. 1982. Mollusca. *In:* Parker, S.P., Synopsis and Classification of Living Organisms, McGraw-Hill, New York, Vol. 1: 945–1166; Vol. 2: 1092–1096.

Chomczynski, P. and N. Scacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry 162: 156–159.

Emberton, K.C, G.S. Kuncio, G.M. Davis, S.M. Phillips, K.M. Monderewicz and Y.H. Guo. 1990. Comparison of recent classifications of stylommatophoran land-snail families, and evaluation of large-ribosomal-RNA sequencing for their phylogenetics. Malacologia 31: 327–352.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27: 401–410.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791.

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. Annual Review in Genetics 22: 521–565.

Felsenstein, J. 1990. Phylip, version 3.3. University of California, Berkeley.

Fretter, V. 1989. The anatomy of some new archaeogastropod limpets (superfamily Peltospiracea) from hydrothermal vents. Journal of Zoology, London 218: 123–169.

Golikov, A.N. and Y.I. Starobogatov. 1988. Problems of phylogeny and system of the prosobranchiate gastropods. *In:* Starobogatov, Y.I. (ed.), Systematics and fauna of Gastropoda, Bivalvia and Cephalopoda. USSR Academy of Sciences, Proceedings of the Zoological Institute 187: 4–77 (in Russian).

Hassouna, N., B. Michot and J.P. Bachellerie. 1984. The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucleic Acids Research 12: 3563–3583.

Haszprunar, G. 1988a. A preliminary phylogenetic analysis of the streptoneurous gastropods. *In:* Ponder, W.F. (ed.),

Prosobranch phylogeny. Malacological Review supplement 4: 7–16.

Haszprunar, G. 1988b. Comparative anatomy of cocculiniform gastropods and its bearing on archeogastropod systematics. *In:* Ponder, W.F. (ed.), Prosobranch phylogeny. Malacological Review, supplement 4: 64–84.

Haszprunar, G. 1988c. On the origin and evolution of major gastropod groups, with special reference to the Streptoneura. Journal of Molluscan Studies 54: 367–441.

Haszprunar, G. 1992. Ultrastructure of the osphradium of the tertiary relict snail, *Campanile symbolicum* Iredale (Mollusca, Streptoneura). Philosophical Transactions of the Royal Society of London B 337: 457–469.

Hickman, C. S., and J.H. McLean. 1990. Systematic revision and suprageneric classification of trochacean gastropods. Natural History Museum of Los Angeles County, Science Series 35: 1–169.

Lecointre , G., H. Philippe, H.L.V. Lê, and H. Le Guyader. 1993. Species sampling has a major impact on phylogenetic inference. Phylogenetics and Molecular Evolution 2: 205–224.

Lecointre, G., H. Philippe, H.L.V. Lê, and H. Le Guyader. In press. How many nucleotides are required to resolve a phylogenetic problem? The practice of a new statistical method on 28 rRNA data. Phylogenetics and Molecular Evolution.

Lindberg, D.R. 1988. The Patellogastropoda. *In:* Ponder, W.F. (ed.), Prosobranch phylogeny. Malacological Review, supplement 4: 35–63.

McLean, J.H. 1981. The Galapagos rift limpet *Neomphalus*: relevance to understanding the evolution of a major Paleozoic-Mesozoic radiation. Malacologia 21: 291–336.

McLean, J.H. 1989. New archaeogastropod limpets from hydrothermal vents: new family Peltospiridae, new superfamily Peltospiracea. Zoologica Scripta 18: 49–66.

McLean, J.H. 1990a. Neolepetopsidae, a new docoglossate limpet family from hydrothermal vents and its relevance to patellogastropod evolution. Journal of Zoology, London 222: 485–528.

McLean, J.H. 1990b. A new genus and species of neomphalid limpet from the Mariana vents with a review of current understanding of relationships among Neomphalacea and Peltospiracea. Nautilus 104: 77–86.

Philippe, H. 1992. MUST, Management Utilitarians for Sequences and Trees. Université Paris Sud, Orsay.

Ponder, W.F. (ed.). 1988. Prosobranch phylogeny. Malacological Review, supplement 4: 1–346.

Ponder, W. F. and A. Warén. 1988. Appendix / Classification of the Caenogastropoda and Heterostropha—A list of the family-group names and higher taxa. *In:* Ponder, W.F. (ed.), Prosobranch phylogeny. Malacological Review, supplement 4: 288–326.

Qu, L.H., B. Michot and J.P. Bachellerie. 1983. Improved methods for structure probing in large RNAs: rapid "heterologous" sequencing approach is coupled to the direct mapping of nuclease accessible sites. Application to the 5′ terminal domain of eukaryotic 28S rRNA. Journal of Molecular Evolution 28: 113–124.

Saitou, N. and M. Nei. 1987. The Neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406–425.

Salwini-Plawen, L. von and G. Haszprunar. 1987. The Vetigastropoda and the systematics of streptoneurous gastropods (Mollusca). Journal of Zoology, London 211: 747–770.

Taylor, J.D. and N.J. Morris. 1988. Relationships of neogastropods. *In:* Ponder, W.F. (ed.), Prosobranch phylogeny. Malacological Review, supplement 4: 167–179.

Tillier, S. 1984. Relationships of gymnomorph gastropods. Zoological Journal of the Linnean Society 82: 345–362.

Tillier, S. 1989. Comparative morphology, phylogeny and classification of land snails and slugs (Gastropoda: Pulmonata: Stylommatophora). Malacologia 31: 1–303.

Tillier, S., M. Masselot, H. Philippe and A. Tillier. 1992. Phylogénie moléculaire des Gastropoda fondée sur le séquençage partiel de l'ARN ribosomique 28S. Comptes-Rendus de l'Académie des Sciences, Paris, série III 314: 79–85.

Warén, A., and P. Bouchet. 1989. New gastropods from East Pacific hydrothermal vents. Zoologica Scripta 18: 67–102.

Wilson, A.C., S.S. Carlsson, and T.J. White. 1977. Biochemical evolution. Annual Review of Biochemistry 46: 573–639.

Site-Directed Mutagenesis with the Polymerase Chain Reaction for Identification of Sibling Species of *Mytilus*

Jonathan B. Geller[1,2]
Dennis A. Powers
Hopkins Marine Station
Stanford University
Pacific Grove, CA 93950 USA

ABSTRACT

The large sample sizes needed for population and biogeographic studies make sequencing impractical as a means of quantifying genotype frequencies. In a previous study, we identified sequence variation in the mitochondrial 16S ribosomal gene from mussels in a sibling species group (*Mytilus* spp.) that discriminated mussels introduced to the west coast of North America (*Mytilus galloprovincialis*) from native mussels (*Mytilus trossulus*). In this study, we used site-directed mutagenesis by the polymerase chain reaction (PCR) to transform a single diagnostic nucleotide substitution into a restriction site. By using a large (51 bp) oligomer as a primer, a size shift in cut and uncut PCR products was visible on an agarose gel. Thus, PCR followed by a restriction digestion allowed identification of species-specific haplotypes in hours rather than days. This method is applicable to any study in which rapid detection of genotypes is desired. We show the presence of *M. trossulus* haplotypes in Monterey Bay, California, extending its southern range. We also present evidence for heteroplasmic individuals that contain mitochondrial haplotypes indicative of both species.

Key Words: Biogeography, genetic introgression, heteroplasmy, mitochondrial DNA, mussels, *Mytilus*, sibling species.

INTRODUCTION

Sibling species are difficult to identify using morphological criteria, perhaps because insufficient time for substantial morphological divergence has elapsed since genetic isolation. Yet, because of the recency of divergence, sibling species are especially important in the study of speciation. Therefore the ability to describe the geographic distribution of sibling species is of considerable interest. Mussels in the genus *Mytilus* are an example of a sibling species group whose identification depends primarily upon genetic analysis. McDonald *et al.* (1991) have summarized the biogeographic evidence provided by analysis of the frequency and distribution of allozymes: *Mytilus edulis* Linné, 1768 occurs primarily on the Atlantic shore of North America and northern Europe. *Mytilus galloprovincialis* Lamarck, 1819 occurs in southern Europe and as introduced populations elsewhere. *Mytilus trossulus* Gould, 1850 occurs on the shores of the northern Pacific Ocean. Our work has focused on the invasion of *Mytilus galloprovincialis* onto the west coast of North America. In this paper, we describe a method for the rapid identification of *M. trossulus* and *M. galloprovincialis* in a mixed population in central California.

Analysis of mitochondrial DNA (mtDNA) polymorphisms within and between populations and closely related species has been used to explore the nature of speciation (Avise, 1986). The earlier investigations of mtDNA variation were based on restriction fragment length polymorphism analysis, requiring purification of mtDNA, restriction digestion, and end-labeling of DNA fragments for visualization on autoradiographs (Hillis & Moritz, 1990). These procedures are time consuming and limit the number of individuals that can be analyzed per population. More recently, the polymerase chain reaction (PCR) and widely applicable universal primers have been used to amplify regions of mtDNA from large numbers of individuals. PCR products can then be analyzed by sequencing or RFLP analysis. Sequencing of large numbers of individuals is usually beyond the scope of population and biogeographic studies, and RFLP analysis of PCR products will not reveal variation in non-restriction sites. Variation in non-restriction sites can be detected by hybridization with oligonucleotide probes. However, reliable selective hybridization typically requires several proximate nucleotide differences, and time consuming radioisotopic labeling, blotting, hybridization, and exposure to autoradiographic film. We devised an alternative approach to detect a single diagnostic base pair difference in the mtDNA genome of the *Mytilus trossulus* and *M. galloprovincialis*.

[1] Current Address: Department of Biological Sciences, University of North Carolina at Wilmington, Wilmington, NC 28403–3297. USA.
[2] Author for correspondence.

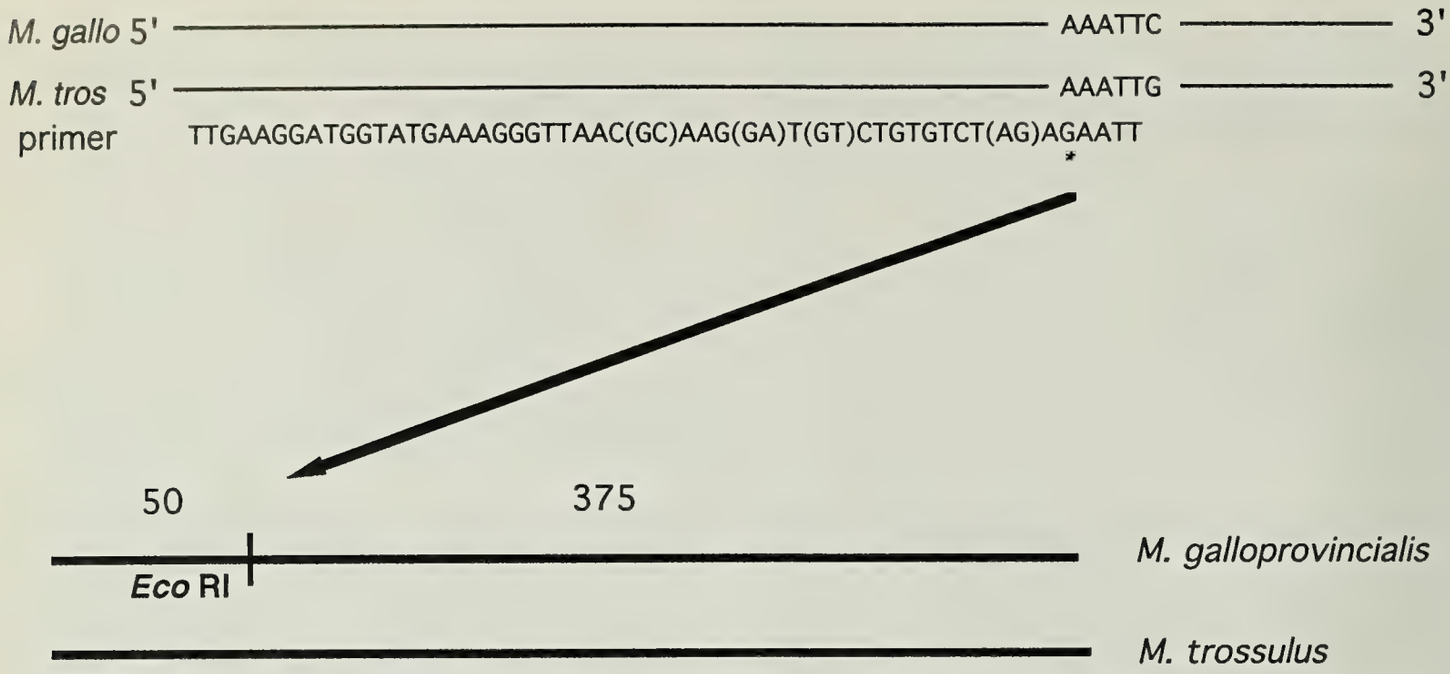


Figure 1. A diagnostic *Eco* RI is created by introducing a mismatch in a PCR primer. Primer MYT16SA-RI changes the sequence AAATTN to GAATTN (asterisk in shown sequence). Only where N is C (*Mytilus galloprovincialis*) is a restriction site produced. *Mytilus galloprovincialis* and *M. trossulus* can thus be distinguished after an *Eco* RI digestion, as illustrated in the bottom panel.

We first used PCR to amplify a portion of the 16S ribosomal RNA gene from the mitochondrial genome of *Mytilus trossulus* and *M. galloprovincialis* (Geller *et al.*, 1993). These products were cloned and sequenced and five diagnostic nucleotide substitutions were found. While none of these substitutions were found within restriction sites, one diagnostic substitution was in a region that differed from an *Eco* RI site by a single base pair in *M. galloprovincialis*, and by two sites in *M. trossulus*. We used site-directed mutagenesis to alter one nucleotide base within this region, producing a restriction site difference for the two species that can be analyzed on agarose gels.

MATERIALS AND METHODS

Site-directed mutagenesis: Site-directed mutagenesis takes advantage of the tolerance of PCR to mismatches in oligonucleotides primers. To convert the site adjacent to a nucleotide substitution into a restriction site, a mutation was introduced in the five nucleotides upstream of the substitution (Figure 1). By changing the sequence AAATTN to GAATTN, an *Eco* RI site is created where N is C. Thus, only *Mytilus galloprovincialis* will have this site. To create a visualizable size shift on an agarose gel after digestion with *Eco* RI, the primer was unusually large (51 bp). The primer was designed to be degenerate at other variable (but not diagnostic) positions to insure annealing to all *Mytilus* haplotypes.

We used mussels that had previously been sequenced as well as mussels of unknown genotype to demonstrate this system of identification. Using the mutagenic primer MYT16A-RI (TTG AAG GAT GGT ATG AAA GGG TTA AC(GC) AAG AAG (GA)T(GT) CTG TGT CT(AG) AGA ATT) and MYT16SB (CCG TTC TGA ACT CAG CTC ATG T), an approximately 425 bp segment of the mitochondrial 16S ribosomal RNA gene was amplified. Each reaction consisted of 2.5 µl 10× reaction buffer (100 mM Tris (pH 8.3), 15 mM $MgCl_2$, 500mM KCl, 1% gelatin, 1% Nonidet–40, and 1% TritonX–100), 20 pmole of each primer, 0.5 units of *Taq* DNA polymerase (Amplitaq, Perkin-Elmer), 0.5 µl of a crude preparation of total cellular DNA, and dH_20 in a total volume of 25 µl. Thirty five cycles of one minute each of denaturation at 94°C, annealing at 54°C, and extension at 72°C were performed on automated thermocyclers (Perkin-Elmer).

Restriction digests: Without purification, 10.3 µl of each PCR product was digested at 37°C for 3–14 hr with 7 units of *Eco* RI using 1.2 µl of the manufacturer's (US Biochemicals) provided 10× buffer. Digests were analyzed on a 4% agarose gel (3% NuSieve, 1% SeaKem, FMC Corp.).

RESULTS

Figure 2 shows the result of mutagenic PCR followed by *Eco* RI restriction digestion for mussels with known genotypes, and demonstrates the effectiveness and ease of scoring of our method for sibling species identification. All amplification products from previously sequenced *Mytilus galloprovincialis*, drawn from Japanese and San Diego populations, contained the mutant *Eco* RI site, while all products from mussels with known *Mytilus trossulus* haplotypes (from Tillamook Bay, Oregon) lacked this restriction site.

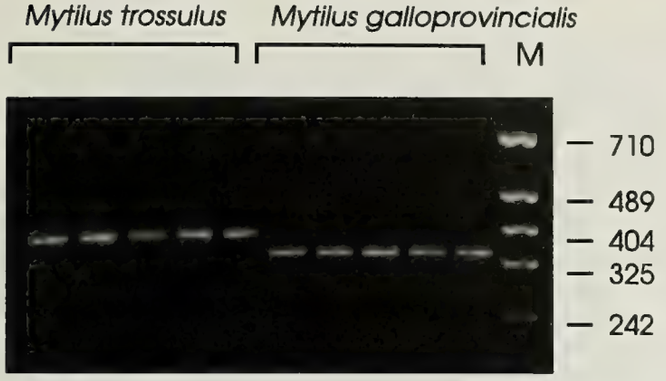


Figure 2. *Eco* RI digests of amplification products generated with primers MYT16SA-RI and MYT16SB. Only products from *Mytilus galloprovincialis* were cut. *Mytilus galloprovincialis* were individuals transported in ballast water as larvae from Japan to Coos Bay, Oregon, where they were cultured to a size sufficient for DNA extraction. *Mytilus trossulus* were from Tillamook Bay, Oregon (see Geller *et al.* 1993 for details). M indicates size marker (*Hpa* II digest of pBluescript KS+).

Figure 3 illustrates the application of this method to an uncharacterized population from Monterey Bay, California. Our analysis showed that this is a mixed population, containing both *Mytilus trossulus* and *M. galloprovincialis*. Some digestions produced both cut and uncut fragments. In these cases, purification of the PCR product followed by digestion with excess enzyme (which had demonstrated activity on other PCR products) did not eliminate the uncut fragment, suggesting the presence of two mitochondrial haplotypes in these mussels. In total, we found 26 *M. trossulus*, 26 *M. galloprovincialis*, and 13 heteroplasmic individuals with haplotypes of both species.

DISCUSSION

Previous studies have used analysis of PCR products to identify sibling or closely related species. For example, Hall and Smith (1991) used restriction digests of amplified mtDNA to distinguish African and European honeybees. In the case of honeybees, restriction sites were found that distinguish subspecies of *Apis mellifera*. An example of an alternative approach is the study of Morin *et al.* (1992), who used oligonucleotide mitochondrial probes to distinguish subspecies of the chimpanzee (*Pan troglodytes*). While useful in these particular studies, both of these methods may not be broadly applicable. For example, PCR products lacking diagnostic restriction sites cannot be distinguished by digestion, and hybridization of oligonucleotide probes is most useful when several substitutions are proximate. Further, many workers may wish to avoid potentially hazardous and time-consuming procedures using radioisotopes. Our method using site-directed mutagenesis only requires that a restriction site can be created by alteration of a site containing a diagnostic substitution.

It is important to note that our species assignments are based on a single nucleotide difference, as assayed by the presence or absence of the mutagenic restriction site. Without prior characterization of sequence variation within and among species considered, undiscovered variation at a "diagnostic" site may lead to misassignment of haplotype or species identity. In the example presented here, we tested the accuracy of this method of identification by applying the assay to populations of mussels from Tillamook Bay, Oregon (n=54) and sites in Japan (n=45) that had been previously characterized by allozyme electrophoresis (McDonald *et al.*, 1991; Wilkins *et al.*, 1981), and found complete agreement between the two techniques (Geller *et al.*, 1994).

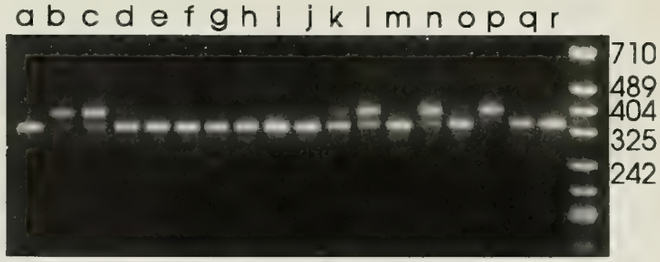


Figure 3. *Eco* RI digests of amplification products from mussels of unknown genotype collected in Monterey Bay, California. Mussels generating uncut products (upper bands; *eg*, lane O) are typed as *Mytilus trossulus*, those generating cut products (lower bands; *eg*, lanes D-J) are typed as *M. galloprovincialis*, and lanes with both upper and lower bands (*eg*, lanes L and N) are from heteroplasmic mussels.

We are applying this method to our study of the geographic distribution of *Mytilus trossulus* and *M. galloprovincialis* on the west coast of North America. Preliminary results are generally in accord with those reported by McDonald and Koehn (1988) in their study of allozyme distribution: *M. trossulus* is prevalent north of San Francisco Bay and *M. galloprovincialis* is most common in southern California. However, the results presented in this paper demonstrate the presence of *M. trossulus* haplotypes in Monterey Bay, south of San Francisco Bay, and work in progress reveal *M. trossulus* haplotypes in southern California. McDonald and Koehn (1988) did not sample in Monterey Bay, and it is possible that their sample of 50 mussels from southern California (25 each from San Diego and Pt. San Luis) simply did not include any *M. trossulus*. Notwithstanding the above, it seems likely that some *M. trossulus* enzyme electromorphs would have been found were they present. We hypothesize that the *M. trossulus* haplotypes we find far south of San Francisco Bay may reflect introgression of *M. trossulus* mtDNA into *M. galloprovincialis* populations. The species making up the *Mytilus edulis* complex are known to hybridize where they overlap in distribution (*eg*, Vainola & Hvilson, 1991), and mtDNA heteroplasmy has been reported several times (Fisher & Skibinski, 1990; Hoeh *et al.*, 1991; Zouros *et al.*, 1992). The presence of heteroplasmic hybrid individuals in Monterey Bay is consistent with our hypothesis.

Our previous study (Geller *et al.*, 1993) of 16S ribosomal rDNA sequences did not reveal differences that

distinguish *Mytilus galloprovincialis* and *M. edulis*. Thus, our assay cannot be used unmodified in localities where these two species coexist. However, once diagnostic nucleotide variation is determined, new mutagenic primers can be designed to apply to studies in these geographic locations. Such studies may be important for resolving the ongoing debate over the systematic status of these two species (Gosling, 1984; Gardner, 1992; McDonald *et al.*, 1991).

ACKNOWLEDGMENTS

We thank J. Rosenthal for suggesting site-directed mutagenesis. We also thank Dr. M.G. Harasewych for inviting JBG to participate in the 11th International Malacology Congress symposium on Molecular Techniques and Molluscan Phylogeny. This work was supported by a fellowship by the National Science Foundation, Division of Ocean Sciences to JBG.

LITERATURE CITED

Avise, J. C. 1986. Mitochondrial DNA and the evolutionary genetics of higher animals. Philosophical Transactions of the Royal Society of London B 312:325–342.

Fisher, C. and D.O.Fisher. 1990. Sex-biased mitochondrial DNA heteroplasmy in the marine mussel *Mytilus*. Proceedings of the Royal Society of London B 242:149–156.

Gardner, J. P. A. 1992. *Mytilus galloprovincialis* (Lmk) (Bivalvia, Mollusca): The taxonomic status of the Mediterranean mussel. Ophelia 35:219–243.

Geller, J. B., J. T. Carlton and D. A. Powers. 1993. Interspecific and intrapopulation variation in mitochondrial ribosomal DNA sequences of *Mytilus* spp. (Bivalvia: Mollusca). Molecular Marine Biology and Biotechnology 2:44–50.

Geller, J. B., J. T. Carlton and D. A. Powers. 1994. PCR-based detection of mtDNA haplotypes of invading and native mussels on the northeastern Pacific coast: latitudinal pattern of invasion and introgression. Marine Biology, in press.

Hall, H. G. and D. R. Smith. 1991. Distinguishing African and European honeybee matrilines using amplified mitochondrial DNA. Proceedings of the National Academy of Science, U.S.A. 88:4548–4552.

Gosling, E. M. 1984. The systematic status of *Mytilus galloprovincialis* in western Europe: a review. Malacologia 25:551–568.

Hillis, D. M. and C. Moritz. 1990. Molecular Systematics. Sinauer Associate, Sunderland, MA 588p.

Hoeh, W. R., K. H. Blakely and W. M. Brown. 1991. Heteroplasmy suggests limited biparental inheritance of *Mytilus* mitochondrial DNA. Science 251:1488–1498.

McDonald, J. H., R. Seed and R. K. Koehn. 1991. Allozyme and morphometric characters of three species of *Mytilus* in the Northern and Southern Hemispheres. Marine Biology 111:323–333.

McDonald, J. H. and R. K. Koehn. 1988. The mussels *Mytilus galloprovincialis* and *M. trossulus* on the Pacific coast of North America. Marine Biology 99: 111–118.

Morin, P. A., J. J. Moore and D. S. Woodruff. 1992. Identification of chimpanzee subspecies with DNA from hair and allele specific probes. Proceedings of the Royal. Society of London B 249:293–297.

Vainola, R. and M. M. Hvilsom. 1991. Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations (Mytilidae: Mollusca). Biological Journal of the Linnean Society. 43:127–148.

Wilkins, N. P., K. Fujino and E. M. Gosling. 1983. The Mediterranean mussel *Mytilus galloprovincialis* Lmk in Japan. Biological Journal of the Linnean Society. 20:265–374

Zouros, E., K. Freeman, A. Oberhauser and G. H. Pogson. 1992. Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus*. Nature 359:412–414.

Morphological and Genetic Variation in Greek Populations of the Edible Snail *Helix aspersa* Müller, 1774 (Gastropoda, Pulmonata): A Preliminary Survey

Maria Lazaridou-Dimitriadou
Y. Karakousis
A. Staikou
Departments of Zoology and Genetics
School of Biology
Faculty of Sciences
Aristotle University of Thessaloniki
54006 Thessaloniki, Macedonia
Greece

ABSTRACT

For the study of genetic and phenotypic variation in allopatric populations of the edible snail *Helix aspersa*, samples were collected (20–40 adult animals/sample) from 24 different regions on the mainland and the islands of Greece as well as Cyprus.

The morphometric data of the shell were analyzed using principal component analysis (PCA). A dendrogram was constructed using UPGMA cluster analysis, based on the Mahalanobis distance between the morphometric parameters of all the populations examined. It showed that the populations from the regions of Peloponesos, the eastern mainland of Greece, Iraklion (island of Crete), and the islands Hios and Cyprus constitute a separate group from the western part of the mainland of Greece and the islands of Crete and Paros.

No relationship was found between altitude and size (D + W) of the animals.

A correlation existed between the size (H/D) of the animals and precipitation ($r = 0.6$, $P < 0.001$) and between D and the mean minimum annual monthly temperatures ($r = 0.4$, $P < 0.001$).

All the populations examined at the isoenzymic level were found to be polymorphic. Three of the examined loci (α GPDH, PMI, SOD) were found to be monomorphic in all populations examined.

The percentage of polymorphic loci (P) ranged from 33.3 to 66.7% and the mean heterozygosity from 0.152 to 0.254.

Two alleles GPD-1B and LAP-1D were found in high frequency in Western Greece but they were absent or in very low frequency in Eastern Greece and the islands.

Key words: Polymorphism, allozymes, *Helix aspersa*.

INTRODUCTION

Helix aspersa Müller, 1774, a terrestrial gastropod with probable origin in the western Mediterranean (Sacchi, 1958) and widely introduced by man to continents (West Europe and North America) and islands of the temperate and tropical zones (Pilsbry, 1939), is one of the more successful colonizing snails. *H. aspersa*'s adaptability is accompanied by an intraspecific variability concerning the polymorphism of the shell, its reproductive system, and its biological cycle (Selander & Kaufman, 1975; Chevallier, 1977; Crook, 1982; Madec & Daguzan, 1987; Madec, 1989; Albuquerque de Matos, 1989; Bleakney *et al.*, 1989).

A visible polymorphism in shell size was found in *H. aspersa* (Lazaridou-Dimitriadou *et al.*, 1983) among populations from two geographical regions of Greece, the island of Crete and Peloponesos. This polymorphism seemed to be related either to differences in the duration of the drought period in the two regions and/or to genetic differentiation. The aim of this study was to determine whether there is intra- and interpopulation variation in morphology and isoenzymes, and to discover the possible correlation between the climatic factors and variation of *H. aspersa* in Greece. Additionally, knowing that shell size can have a genetic component in helicid and other snails (Cook, 1965; Baur, 1984; Goodfriend, 1986), we tried to determine if the morphological differentiation of populations coming from the islands of Crete and Peloponesos also manifests itself in biochemical polymorphism. If this is so, we have an apparent case of White's "geographical races" (White, 1978) and the existence of two forms, *H. aspersa aspersa* and *H. aspersa major* (Chevallier, 1977).

METHODS AND MATERIALS

Sampling: We sampled 24 populations of *H. aspersa*, totalling 805 adult individuals (3 populations from the northwestern part of Greece, 5 from Central Greece, 7 from Peloponesos, 8 from the Aegean islands and one

Figure 1. Sites of samples in Greece and Cyprus.

from Cyprus island) (average sample size per population = 30) (figure 1).

In each sample, the specimens were taken in an area of approximately 400 m^2. The meteorological data were obtained from the Data Department of National Meteorological Institute of Greece.

Living specimens were brought to the laboratory in carton boxes, where they were aestivating.

Morphometric data: Morphometric and qualitative data of the shell were taken immediately before electrophoretic analysis.

The snails were weighed (W) and individually marked on the shell. Maximum shell diameter (D), aperture diameter (di), shell height (H) and shell thickness (T), were measured with a digital calliper to the nearest of 0.01 mm. The aperture area (Ap) was copied on tracing paper and measured with a planimeter. The color was recorded according to Albuquerque de Matos' system (1984) and number and fusion of bands was also noted. Four classes existed for colour (yellow, brown, red, white), six for the number of bands (0, 1, 2, 3, 4, 5) and 13 for the fusion of bands (observed combination of fusion of the existing bands).

Electrophoresis: For the assessment of the genetic polymorphism, 13 enzymatic systems were investigated using starch gel electrophoresis. These enzymatic systems correspond to 15 genetic loci. Buffers and tissues used are shown in table 1. The alleles were named (A–E) from greater to less mobility. Several esterase loci were present, but only two loci with zymograms interpretable in Mendelian terms were used.

The samples for the electrophoretic analysis were prepared from the foot muscle and hepatopancreas (table 1) (after the complete removal of the intestine and washing in distilled water in order to remove any parasites (Haralambidis *et al.*, 1985). Organs were then finely cut and homogenized in equal volume of 0.1 N $CaCl_2$-1.7% sucrose and stored at −20 °C for up to 15 days.

Statistical analysis: In order to look for geographical variation, the morphometric and qualitative data of the shells were subjected to principal component analysis (PCA) using the Statview computer program for Macintosh. The principal component analyses (PCA) were done using either the morphometric characters or their ratios (because H was highly correlated with D, di with D, di with H and Ap with D): the maximum shell diameter (D), the weight (W), the thickness of the shell (T) and the ratios of height/maximum shell diameter (H/D), of (di)/D, of di/H, and of D/Ap. Transformations to natural logarithms were used for D, W, T and arcsin transformations for the ratios. To compare the scores of the first 3 factors of the PCA, ANOVA test was also made using Statview computer program for Macintosh.

To determine relationships, the morphometric variables were correlated with the climatic data. Transformations to standard scores were used for the climatic

Table 1. Electrophoretic conditions and enzymes analyzed.

Enzyme	Abbrev.	E.C. no.	Locus	Tissue	Buffer
Aspartate aminotransferase	AAT	2.6.1.1.	1	Hep.	A
Esterase	EST	3.1.1.	2	Hep.	B
Esterase D	ESD	3.1.1.	1	Hep.	A
α-Glycerophosphate dehydrogenase	αGPDH	1.2.1.12.	1	Foot	A
Glucosephosphate isomerase	GPI	5.3.1.9.	1	Hep.	C
Isocitrate dehydrogenase	IDH	1.1.1.42.	1	Foot	A
Lactate dehydrogenase	LDH	1.1.1.27.	1	Hep.	B
Leucine aminopeptidase	LAP	3.4.11.–	1	Hep.	C
Malate dehydrogenase	MDH	1.1.1.37.	1	Hep.	A
Phosphomannose isomerase	PMI	5.3.1.8.	1	Hep.	C
6-phosphogluconate dehydrogenase	GPD	1.1.1.49.	1	Foot	A
Phosphoglucomutase	PGM	2.7.5.1.	2	Hep.	C
Superoxide dismutase	SOD	1.15.1.1.	1	Hep.	B

A = 0.13 M Tris, 0.04 M citric acid, 0.018 M EDTA, pH = 7.1 for gel buffer; 0.13 M Tris, 0.037 M Citric acid, 0.001 M EDTA, pH = 7.1 electr. buffer. B = Ashton *et al.* (1961). C = Smith (1976).

Table 2. The mean yearly meteorological data over 10 years (1980–1990). T = mean maximum annual monthly temperature, Tmi = mean minimum annual monthly temperature, Tmax = maximum annual monthly temperature, Tmin = minimum annual monthly temperature, Prec. = total annual precipitation in mm. Mean annual drought duration in days (estimated from the drought period of the ombrothermic curve of each region from 1986 to 1990).

Localities	T (°C)	Tmi	Tmax	Tmin	Prec. (mm)	Drought duration (in days)
Igoumenitsa	26.98	9.32	37.32	−2.35	975	120
Igoumen.—2	26.98	9.32	37.32	−2.35	975	120
Preveza	25.72	8.66	35.2	0.036	850	153
Mesolongi	27.98	9.07	37.21	−0.29	373	270
Nafpaktos	28.85	9.85	37.88	−2.53	726	129
Athens	27.98	9.07	37.21	−0.29	373	114
Halkida	28.21	8.39	39.9	−1.92	376	207
Karystos	27.03	9.66	35.35	0.59	666	198
Lamia	25.86	6.08	40.08	—	597.	162
Kam. Vourla	25.86	6.08	40.08	—	597	162
Paros	25.09	10.3	35.57	1.3	423	207
Kiato	27.76	8.01	39.01	−0.61	476	165
Nafplio	26.85	8.33	38.43	−1.1	527	150
Glykovrysi	27.12	9.84	38.4	−0.28	836.	180
Karavas	28.88	8.88	41.38	− 1.74	915	162
Kyparissia	25.61	10.14	35.14	0.23	762	156
Zaharo	26.63	8.96	38.13	−3.23	896	156
Karytaina	24.9	4.55	39.28	−10.12	788	126
Kandanos	28.25	11.94	39.53	3.78	311	243
Tympaki	26.21	10.63	39.91	1.43	514	192
Siteia	25.82	11.16	35.13	3.23	520	171
Irakleio	26.12	11.02	36.33	2.5	505	150
Hios	26.56	8.98	36.1	−1.5	522	203
Cyprus	28.3	10.1	35.5	7.65	351	210

data. The meteorological data, namely total annual precipitation (P) in mm, absolute maximum annual monthly temperatures (Tmax), absolute minimum annual monthly temperatures (Tmin), mean maximum annual monthly temperatures (T), mean minimum annual monthly temperatures (Tmi), were obtained from the Data Department of National Meterological Institute of Greece (table 2). The mean number of days of monthly drought period (Dr) and the mean number of days of monthly rainfall > 100 mm (Mr) were calculated from the ombrothermic yearly curves.

Seven morphometric variables were used for the Mahalanobis distance. UPGMA cluster analysis (Sneath & Sokal, 1963) was performed on the basis of Mahalanobis distances (1936) using the morphometric data for the 24 populations.

Observed or expected heterozygosity, proportion of polymorphic loci for each population, mean alleles per locus, genetic distances (Nei, 1978) and F-statistics (Wright, 1978) between populations were calculated using BIOSYS-1 (Swofford & Selander, 1981).

To investigate the genetic relationship among the examined populations, a dendrogram based on allelic frequencies, using the complete linkage method based on Nei's genetic distance and a WAGNER tree, were constructed with the aid of BIOSYS-1 computer program.

To compare the degree of differentiation within and between populations, hierarchical F-statistics analysis (Wright, 1978) was performed using BIOSYS-1 computer program.

Number of alleles per locus, the percentage of polymorphic loci, the observed (Ho) and expected heterozygosity (He), and allele frequencies were each correlated with the climatic characteristics of each region.

A correlation analysis was performed between the Nei's genetic distance and the Mahalanobis distance.

RESULTS

I. Morphological Polymorphism

Interpopulational size variation was conspicuous. The mean size of the largest population was statistically different ($P < 0.01$), nearly ⅓ to ¼ larger than the mean size of the smallest population. On the other hand, size variation within populations was far less conspicuous (table 3).

Shells were usually brown-yellowish, and, in most cases, with 5 dark bands, apart from those from Preveza. Some populations had less than 5 bands (3–14%), while in the majority of populations some of the bands were fused.

The first (D + W) (size component) and the second factor (H/D) (elevation component) of PC analysis distinguished Siteia, Hios, Cyprus and the regions of Peloponesos, Kamena Vourla, Lamia, Halkida, Paros and Irakleio as separate groups (table 4, figure 2), but ANO-

Table 3. Means ± standard errors of the studied shell characters in the 24 representative samples of *Helix aspersa* from Greece and Cyprus. Where D: shell largest diameter, W: snail weight, H: shell height, T: shell thickness, di: aperture diameter, Ap.: aperture area, se: standard error.

Regions	D ± se (mm)	W ± se (g)	H ± se (mm)	T ± se (mm)	di ± se	Ap ± se	H/D ± se	% banded	Individuals with <5 bands
Igoumenitsa	31.428 ± 0.297	5.679 ± 0.16	21.403 ± 0.248	0.299 ± 0.012	18.643 ± 0.195	2.994 ± 0.079	0.682 ± 0.008	100	0
Igoumen.—2	29.515 ± 0.31	4.6 ± 0.204	20.85 ± 0.288	0.26 ± 0.017	17.246 ± 0.162	2.556 ± 0.07	0.707 ± 0.007	100	0
Preveza	32.389 ± 0.434	7.057 ± 0.281	23.329 ± 0.345	0.318 ± 0.011	18.904 ± 0.219	3.117 ± 0.082	0.721 ± 0.005	53	14
Mesolongi	33.244 ± 0.339	7.388 ± 0.276	22.737 ± 0.329	0.36 ± 0.016	19.142 ± 0.187	3.103 ± 0.074	0.685 ± 0.009	100	0
Nafpaktos	32.109 ± 0.271	6.467 ± 0.223	21.326 ± 0.291	0.306 ± 0.017	18.236 ± 0.17	2.951 ± 0.071	0.664 ± 0.007	100	0
Athens	34.556 ± 0.467	6.111 ± 0.305	23.427 ± 0.378	0.322 ± 0.019	20.135 ± 0.636	3.137 ± 0.124	0.679 ± 0.009	100	0
Halkida	37.139 ± 0.406	8.388 ± 0.326	24.769 ± 0.353	0.372 ± 0.015	21.428 ± 0.199	3.986 ± 0.102	0.668 ± 0.009	100	0
Karystos	34.023 ± 0.317	6.012 ± 0.206	22.885 ± 0.256	0.272 ± 0.013	19.722 ± 0.219	3.417 ± 0.076	0.674 ± 0.008	100	0
Lamia	36.279 ± 0.371	7.747 ± 0.21	24.537 ± 0.297	0.388 ± 0.018	20.912 ± 0.195	3.592 ± 0.084	0.677 ± 0.007	100	3
Kam. Vourla	38.691 ± 0.381	9.145 ± 0.343	26.26 ± 0.337	0.341 ± 0.019	21.682 ± 0.194	4.279 ± 0.11	0.679 ± 0.008	100	0
Paros	33.033 ± 0.336	8.515 ± 0.332	23.309 ± 0.249	0.48 ± 0.021	19.306 ± 0.341	3.051 ± 0.084	0.706 ± 0.006	100	3
Kiato	36.702 ± 0.47	7.895 ± 0.486	24.638 ± 0.432	0.323 ± 0.018	21.065 ± 0.245	3.871 ± 0.11	0.671 ± 0.007	100	3
Nafplio	37.829 ± 0.469	9.288 ± 0.439	26.614 ± 0.461	0.388 ± 0.017	21.239 ± 0.264	4.017 ± 0.102	0.703 ± 0.007	100	9
Glykovrysi	35.906 ± 0.414	8.323 ± 0.295	24.632 ± 0.402	0.356 ± 0.014	21.147 ± 0.255	3.797 ± 0.116	0.686 ± 0.008	100	0
Karavas	36.809 ± 0.51	7.9 ± 0.394	24.59 ± 0.398	0.295 ± 0.012	20.924 ± 261	3.842 ± 0.129	0.869 ± 0.008	100	0
Kyparissia	37.078 ± 0.39	10.253 ± 0.268	26.093 ± 0.374	0.429 ± 0.011	21.366 ± 0.271	3.919 ± 0.101	0.704 ± 0.009	100	3
Zaharo	38.869 ± 0.762	10.391 ± 0.731	26.689 ± 0.479	0.345 ± 0.017	23.014 ± 0.313	4.523 ± 0.135	0.688 ± 0.007	100	0
Karytaina	38.185 ± 0.46	9.462 ± 0.478	26.256 ± 0.397	0.382 ± 0.017	22.164 ± 0.294	4.236 ± 0.104	0.688 ± 0.009	100	0
Kandanos	32.708 ± 0.414	6.417 ± 0.306	22.459 ± 0.357	0.266 ± 0.017	19.361 ± 0.227	3.129 ± 0.11	0.687 ± 0.007	100	0
Tympaki	32.489 ± 0.427	6.341 ± 0.225	22.596 ± 0.33	0.333 ± 0.015	19.202 ± 0.268	3.172 ± 0.116	0.696 ± 0.007	100	0
Siteia	29.37 ± 0.279	4.706 ± 0.163	19.974 ± 0.211	0.377 ± 0.015	17.416 ± 0.16	2.475 ± 0.055	0.681 ± 0.006	100	0
Irakleio	32.626 ± 0.327	8.997 ± 0.314	22.449 ± 0.309	0.345 ± 0.021	19.689 ± 0.173	3.189 ± 0.062	0.688 ± 0.008	100	0
Hios	34.727 ± 0.529	9.749 ± 0.338	22.847 ± 0.375	0.315 ± 0.023	20.387 ± 0.304	3.49 ± 0.11	0.659 ± 0.009	100	0
Cyprus	37.231 ± 0.366	9.873 ± 0.357	23.107 ± 0.264	0.359 ± 0.017	21.098 ± 202	3.926 ± 0.103	0.621 ± 0.006	100	0

Table 4. Measures of Variable Sampling Adequacy in Principal Component analysis (Total matrix sampling adequacy: 0.495) and the eigenvalues and proportion of original variance in PCA using only morphometric parameters after logarithmic transformation.

Variable	Adeq.	Factor	Magnitude	Variance proportion
H	0.258	1	1.929	0.321
D	0.179	2	1.503	0.251
W	0.384	3	1.436	0.239
T	0.173	4	1.123	0.187
di	0.149	5	0.005	0.001
Ap	0.156			

Bartlett Test of Sphericity—DF: 20, Chi Square = 7470.749, $P < 0.0001$.

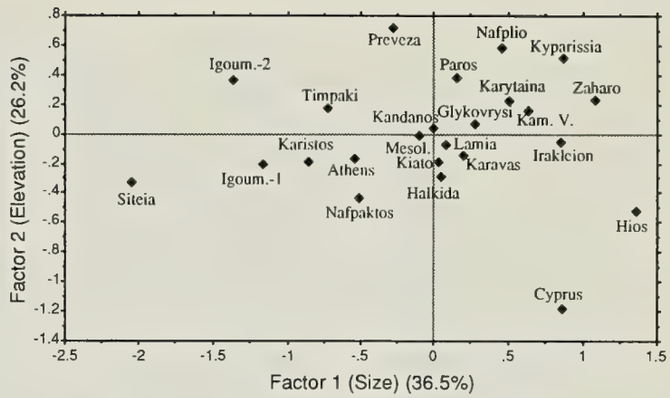


Figure 2. Distribution of populations on Factor I and II scores of principal component analysis based on morphometric data of the shell characteristics in *Helix aspersa*.

VA and Fischer LSD tests on the first three factors of PCA showed statistically significant differences ($P < 0.01$) only in the size of the snails between the populations from Peloponesos and that from Siteia.

The dendrogram of UPGMA cluster analysis based on Mahalanobis' distances (figure 3) was consistent with PCA results: the populations from the regions of Peloponesos and Hios, Irakleio in Crete and Cyprus island and regions from the SE Greece (Kamena Vourla, Lamia, Halkida) constitute a separate group from Crete, NW regions of Greece and Athens, Paros island and Karystos (Evia island).

A correlation was found between the shell thickness, the H/D and the annual precipitation ($r = -0.365$, $P < 0.01$; $r = 0.632$, $P < 0.001$, respectively), between D, the aperture area and the mean minimum annual monthly temperatures ($r = -0.422$, $P < 0.001$; $r = -0.509$, $P < 0.001$, respectively).

II. Biochemical Polymorphism

A high degree of genetic variation was found in all populations examined. Three loci (AGP, PMI, SOD) were monomorphic in all populations examined, while the remaining loci had up to 5 alleles (LAP-1 and EST-2) (table 5).

The percentage of polymorphic loci ranged from 33.3 to 66.7% and the mean expected heterozygosity per locus from 0.152 to 0.254.

Two alleles, GPD-1B and LAP-1D, were found in high frequency in western Greece but were absent or occurred in very low frequency in Eastern Greece and the islands. A unique allele in high frequency in the locus PGM-2 was found in the population of Karytaina.

Coefficients of genetic identity and distance (Nei, 1978) were calculated for all pairwise combinations of the 24 populations studied.

The mean values of genetic distance (D) between all populations was 0.038 ± 0.002 ($N = 276$) and ranged from 0.001 to 0.131. Estimates of Hierarchical F statistics analysis are shown in table 3. There is considerable differentiation at all levels, but the significant point is that the variance among populations of blocks within demes (F_{DR}) was greater than that between regions and the total (F_{RS}, F_{RT}) (table 6). It must be pointed out that GPD had greater values of differentiation in the higher levels (table 6). Average genetic distances within regions used in hierarchical F statistics analysis are shown in table 7. It is obvious that populations coming from northwestern Greece showed a degree of genetic distance from the rest of the populations. This is also shown by a dendrogram constructed using complete linkage method and by the WAGNER tree (figures 4, 5).

To determine whether climatic parameters were correlated with allozymic variation, five climatic variables were chosen, namely total annual precipitation (P), mean maximum annual monthly temperatures (T), absolute maximum annual monthly temperatures (Tmax), mean annual drought period in days (DR) (from the ombrothermic curves) and mean annual rainfall duration >

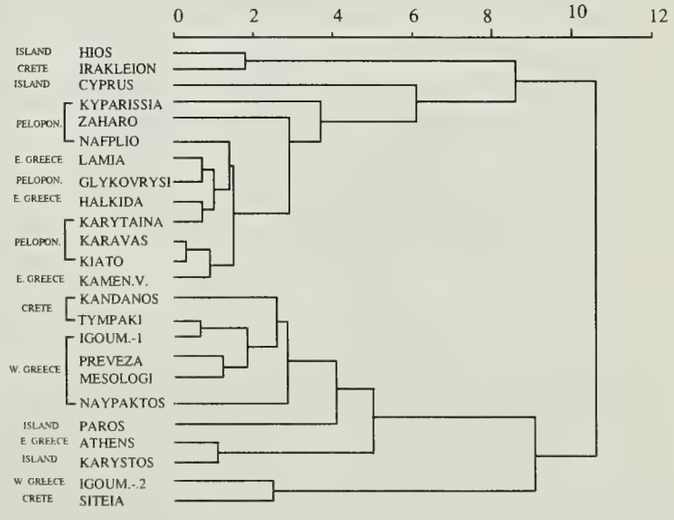


Figure 3. UPGMA cluster analysis based on Mahalanobis distances between 24 populations of *Helix aspersa* using morphometric data.

Table 5. Allele frequencies at twelve enzymatic loci in 24 populations of *Helix aspersa* from Greece and Cyprus.

Locus	Alleles	Igoum.	Igoum.—2	Preveza	Mesol.	Nafpak.	Athens	Halk.	Karyst
AAT-1	A	1.000	1.000	1.000	1.000	1.000	1.000	0.944	1.000
	B	0.000	0.000	0.000	0.000	0.000	0.000	0.056	0.000
EST-D	A	0.125	0.250	0.319	0.569	0.653	0.708	0.681	0.444
	B	0.083	0.208	0.306	0.306	0.208	0.236	0.236	0.444
	C	0.417	0.250	0.181	0.014	0.042	0.056	0.056	0.042
	D	0.375	0.292	0.194	0.111	0.097	0.000	0.028	0.069
EST-1	A	0.986	0.958	0.972	0.958	0.931	0.847	0.931	0.694
	B	0.014	0.042	0.014	0.042	0.042	0.125	0.069	0.167
	C	0.000	0.000	0.014	0.000	0.028	0.028	0.000	0.139
	D	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EST-2	A	0.139	0.347	0.236	0.333	0.347	0.236	0.208	0.083
	B	0.444	0.319	0.222	0.347	0.347	0.222	0.306	0.333
	C	0.194	0.222	0.389	0.139	0.069	0.194	0.194	0.375
	D	0.139	0.083	0.153	0.125	0.194	0.167	0.278	0.167
	E	0.083	0.028	0.000	0.056	0.042	0.181	0.014	0.042
GPD-1	A	0.167	0.153	0.264	1.000	1.000	0.986	1.000	1.000
	B	0.653	0.833	0.639	0.000	0.000	0.014	0.000	0.000
	C	0.181	0.014	0.097	0.000	0.000	0.000	0.000	0.000
IDH-1	A	1.000	1.000	1.000	1.000	0.875	1.000	1.000	1.000
	B	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000
	C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LAP-1	A	0.417	0.403	0.375	0.917	0.708	0.458	0.611	0.958
	B	0.000	0.000	0.014	0.056	0.028	0.153	0.167	0.000
	C	0.097	0.089	0.056	0.028	0.125	0.292	0.111	0.028
	D	0.403	0.375	0.472	0.000	0.139	0.042	0.000	0.014
	E	0.083	0.153	0.083	0.000	0.000	0.056	0.111	0.000
LDH-1	A	0.972	1.000	0.847	0.931	0.944	0.944	0.861	0.917
	B	0.028	0.000	0.153	0.069	0.056	0.056	0.139	0.083
	C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	D	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MDH-1	A	0.347	0.417	0.375	0.441	0.529	0.542	0.403	0.444
	B	0.333	0.306	0.278	0.441	0.265	0.208	0.375	0.319
	C	0.264	0.014	0.236	0.029	0.088	0.056	0.125	0.125
	D	0.056	0.264	0.111	0.088	0.118	0.194	0.097	0.111
PGM-1	A	0.889	0.958	0.986	0.944	0.944	0.931	0.917	0.972
	B	0.056	0.028	0.000	0.028	0.056	0.056	0.000	0.028
	C	0.056	0.014	0.014	0.028	0.000	0.014	0.083	0.000
PGM-2	A	0.889	0.944	0.958	0.972	0.972	1.000	0.819	0.917
	B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	C	0.111	0.056	0.028	0.028	0.028	0.000	0.167	0.083
	D	0.000	0.000	0.014	0.000	0.000	0.000	0.014	0.000
PGI-1	A	0.986	0.944	0.986	0.931	0.931	0.889	0.917	1.000
	B	0.000	0.028	0.000	0.069	0.000	0.000	0.000	0.000
	C	0.000	0.000	0.014	0.000	0.042	0.028	0.042	0.000
	D	0.014	0.028	0.000	0.000	0.028	0.083	0.042	0.000
N		36	36	36	36	36	36	36	36
Percent.		66.67	60.00	66.67	60.00	66.67	60.00	66.67	53.33
H		.253 ± .079	.233 ± .079	.254 ± .082	.173 ± .065	.208 ± .065	.221 ± .073	.240 ± .068	.196 ± .071

100 mm. Results showed that mean observed heterozygosity was negatively correlated with drought duration ($r = -0.419$, $P < 0.05$). Individual alleles also gave significant correlations with P at the level of $P < 0.01$ (EST-DA ($r = -0.647$), EST-DC ($r = 0.581$), EST-DD ($r = 0.797$), GPD-1A ($r = -0.753$), GPD-1B ($r = 0.762$), LAP-1B ($r = -0.506$), LAP-1D ($r = 0.505$), and as for drought LAP-1A ($r = 0.498$)).

DISCUSSION

The study of shell morphometric characteristics of *Helix aspersa* showed an interpopulation variation, but not any trend in geographic variations of the seven variable characters studied jointly. However, there seemed to exist a noticeable cline in some of the characters such as the thickness of the shell, peristome diameter, maximum

Table 5. Extended.

Lamia	K. Vour	Paros	Kiato	Nafplio	Glykov.	Karavas	Kyparis.	Zaharo
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.625	0.833	0.347	0.444	0.694	0.389	0.306	0.319	0.333
0.250	0.069	0.611	0.486	0.208	0.236	0.042	0.069	0.193
0.028	0.083	0.000	0.000	0.000	0.208	0.431	0.389	0.283
0.097	0.014	0.042	0.069	0.097	0.167	0.222	0.222	0.200
0.833	0.861	0.667	0.806	0.875	0.986	0.972	0.875	0.967
0.097	0.069	0.292	0.181	0.028	0.000	0.000	0.083	0.033
0.069	0.028	0.042	0.000	0.042	0.014	0.028	0.042	0.000
0.000	0.042	0.000	0.014	0.056	0.000	0.000	0.000	0.000
0.347	0.181	0.222	0.264	0.181	0.514	0.639	0.208	0.333
0.222	0.417	0.458	0.264	0.292	0.264	0.028	0.278	0.300
0.250	0.153	0.167	0.222	0.292	0.069	0.111	0.319	0.150
0.167	0.222	0.125	0.069	0.181	0.111	0.111	0.181	0.217
0.014	0.028	0.028	0.181	0.056	0.042	0.111	0.014	0.000
1.000	0.917	0.986	1.000	1.000	0.569	0.792	0.944	0.550
0.000	0.042	0.014	0.000	0.000	0.431	0.208	0.056	0.450
0.000	0.042	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	1.000	0.972	0.986	0.972	0.958	1.000	1.000	1.000
0.000	0.000	0.028	0.014	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.028	0.042	0.000	0.000	0.000
0.653	0.639	0.903	0.875	0.708	0.917	0.944	0.708	0.817
0.111	0.083	0.000	0.014	0.042	0.042	0.000	0.000	0.000
0.139	0.194	0.069	0.097	0.139	0.000	0.042	0.292	0.183
0.014	0.028	0.028	0.014	0.083	0.000	0.000	0.000	0.000
0.083	0.056	0.000	0.000	0.028	0.042	0.014	0.000	0.000
0.875	0.931	0.931	0.917	0.903	0.972	0.972	0.986	0.983
0.125	0.069	0.028	0.083	0.097	0.028	0.028	0.014	0.017
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.042	0.000	0.000	0.000	0.000	0.000	0.000
0.389	0.333	0.625	0.403	0.333	0.319	0.194	0.389	0.333
0.431	0.417	0.208	0.389	0.403	0.264	0.444	0.236	0.417
0.111	0.167	0.056	0.083	0.139	0.361	0.097	0.167	0.183
0.069	0.083	0.111	0.125	0.125	0.056	0.264	0.208	0.067
0.931	0.819	0.944	0.917	0.917	0.889	0.931	0.889	0.967
0.056	0.139	0.056	0.083	0.028	0.111	0.056	0.083	0.000
0.014	0.042	0.000	0.000	0.056	0.000	0.014	0.028	0.033
1.000	0.889	1.000	1.000	0.889	1.000	0.972	0.972	1.000
0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.083	0.000	0.000	0.111	0.000	0.028	0.028	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.972	0.889	0.917	0.944	0.944	0.944	0.903	0.903	0.967
0.000	0.083	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.028	0.069	0.014	0.056	0.056	0.097	0.097	0.000
0.028	0.000	0.014	0.042	0.000	0.000	0.000	0.000	0.033
36	36	36	36	36	36	36	36	30
53.33	66.67	66.67	60.00	66.67	66.67	66.67	66.67	60.00
.213 ± .071	.236 ± .064	.194 ± .063	.202 ± .069	.233 ± .067	.215 ± .073	.190 ± .065	.227 ± .074	.214 ± .077

shell diameter and the ratio H/D along a gradient in climatic conditions. The thickness of the shell was positively correlated with low precipitation as also shown by Bar (1978) in *Theba pisana*, and the height of the shell with high precipitation as mentioned by Goodfriend (1986). In our case it seemed that aperture area and the maximum shell diameter were negatively related to the absolute minimum annual monthly temperature, suggesting that larger snails survive better in colder conditions, and that conditions of the environment might select for size in land snails through differential mortality (Goodfriend, 1986).

Additionally, a great variability between populations was found at the biochemical level. However, there did not exist any correlation between Mahalanobis and Nei's genetic distance of the 24 populations studied, and the

Table 5. Extended.

Locus	Alleles	Karyt.	Kand.	Tympaki	Siteia	Hios	Cypru	Iraklei
AAT-1	A	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	B	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EST-D	A	0.222	0.681	0.792	0.552	0.450	0.371	0.743
	B	0.222	0.083	0.028	0.086	0.450	0.143	0.186
	C	0.431	0.194	0.097	0.224	0.075	0.357	0.071
	D	0.125	0.042	0.083	0.138	0.025	0.129	0.000
EST-1	A	0.958	0.847	0.764	0.879	0.525	1.000	1.000
	B	0.014	0.111	0.194	0.086	0.150	0.000	0.000
	C	0.014	0.028	0.028	0.034	0.325	0.000	0.000
	D	0.014	0.014	0.014	0.000	0.000	0.000	0.000
EST-2	A	0.611	0.444	0.458	0.362	0.400	0.329	0.671
	B	0.042	0.181	0.125	0.121	0.300	0.186	0.014
	C	0.028	0.083	0.083	0.190	0.000	0.114	0.271
	D	0.181	0.139	0.333	0.103	0.100	0.329	0.043
	E	0.139	0.153	0.000	0.224	0.200	0.043	0.000
GPD-1	A	1.000	1.000	1.000	0.966	0.875	1.000	0.929
	B	0.000	0.000	0.000	0.000	0.125	0.000	0.000
	C	0.000	0.000	0.000	0.034	0.000	0.000	0.071
IDH-1	A	1.000	1.000	1.000	1.000	0.850	1.000	1.000
	B	0.000	0.000	0.000	0.000	0.150	0.000	0.000
	C	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LAP-1	A	0.861	0.708	0.500	0.776	0.825	0.957	0.586
	B	0.000	0.083	0.043	0.017	0.000	0.000	0.029
	C	0.139	0.125	0.186	0.052	0.175	0.000	0.343
	D	0.000	0.028	0.086	0.052	0.000	0.043	0.043
	E	0.000	0.056	0.186	0.103	0.000	0.000	0.000
LDH-1	A	1.000	0.667	0.889	0.948	1.000	1.000	1.000
	B	0.000	0.333	0.111	0.052	0.000	0.000	0.000
	C	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	D	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MDH-1	A	0.338	0.472	0.333	0.207	0.275	0.257	0.300
	B	0.250	0.222	0.431	0.379	0.275	0.243	0.300
	C	0.294	0.153	0.181	0.328	0.375	0.329	0.057
	D	0.118	0.153	0.056	0.086	0.075	0.171	0.343
PGM-1	A	0.833	0.917	0.917	0.897	1.000	0.986	1.000
	B	0.167	0.083	0.056	0.103	0.000	0.000	0.000
	C	0.000	0.000	0.028	0.000	0.000	0.014	0.000
PGM-2	A	0.278	1.000	0.972	1.000	1.000	0.971	1.000
	B	0.500	0.000	0.000	0.000	0.000	0.000	0.000
	C	0.000	0.000	0.000	0.000	0.000	0.029	0.000
	D	0.222	0.000	0.028	0.000	0.000	0.000	0.000
PGI-1	A	1.000	0.833	0.958	0.966	1.000	0.986	0.000
	B	0.000	0.042	0.000	0.017	0.000	0.000	0.000
	C	0.000	0.069	0.028	0.000	0.000	0.014	0.000
	D	0.000	0.056	0.014	0.017	0.000	0.000	0.000
N		36	36	36	29	20	35	35
Percent.		46.67	53.33	60.00	60.00	46.67	46.67	33.33
H		.217 ± .076	.238 ± .07	.218 ± .069	.210 ± .072	.229 ± .07	.160 ± .077	.152 ± .064

degree of heterozygosity and the coefficient of variation of the maximum shell diameter (D), shell height (H), and shell thickness (T). The inconsistency between UPGMA cluster analysis based on the Mahalanobis distances and the dendrogram found by cluster analysis of complete linkage method using Nei's identity, could be the result of a size effect. If size correction is not used, it seems that intracluster size variation masks intercluster variation (Lazaridou-Dimitriadou *et al.*, in press).

A unique allele (PGM-2B) that was found only in Karytaina's population in high frequency may be due either to founder effect or to altitude. A similar report for locus PGM2 in Spanish *C. nemoralis* (Mazon *et al.*, 1988) attributed this to altitude. Since Karytaina is the only locality that had 555 m elevation, altitude may be the causal factor in *H. aspersa*.

The fact that two alleles, GPD-1B and LAP-1D, were found in high frequencies in Western Greece may be

Table 6. Wright's hierarchical analysis between demes (D), subdivisions (S), and the total area (T).

	F_{DR}	F_{DS}	F_{DT}	F_{RS}	F_{RT}	F_{ST}
AAT	−0.004	−0.009	−0.010	−0.005	−0.006	−0.001
EST-D	−0.004	−0.009	−0.010	−0.005	−0.006	−0.001
EST-1	0.081	0.059	0.047	−0.024	−0.037	−0.013
EST-2	0.027	0.011	0.009	−0.016	−0.018	−0.002
GPD	0.074	0.014	0.383	−0.065	0.333	0.374
IDH	0.065	0.018	0.016	−0.050	−0.053	−0.002
LAP-1	0.029	0.014	0.076	−0.015	0.048	0.062
LDH	0.021	0.005	−0.001	−0.017	−0.023	−0.006
MDH	0.008	0.003	−0.001	−0.005	−0.009	−0.004
PGI	−0.006	−0.006	−0.005	−0.000	0.000	0.000
PGM-1	−0.003	0.001	−0.002	0.004	0.001	−0.003
PGM-2	−0.045	−0.037	−0.052	0.008	−0.007	−0.014
Total	0.044	0.017	0.053	−0.028	0.010	0.037

related to climatic factors (*e.g.*, precipitation). Since their frequencies were found to be significantly correlated to mean annual precipitation, selective forces may be acting on those alleles. However, a positive correlation between the geographic distance of the continental populations and the Nei's genetic distance was found, and stochastic processes cannot be excluded. Madec (1991) also reported a similar change in two loci (MDH-1, PGM-s) from one locality to the other. The same phenomenon has been observed for *C. nemoralis* (Johnson, 1976) and *H. aspersa* (Selander & Ochman, 1983).

The results of hierarchical F statistics showed that heterogeneity is considerable at the intrapopulation level as also reported by Selander and Ochman (1983) for *H. aspersa* in California. However, the heterogeneity of F_{DR} does not seem to support random genetic drift as the cause for the genetic differentiation between localities.

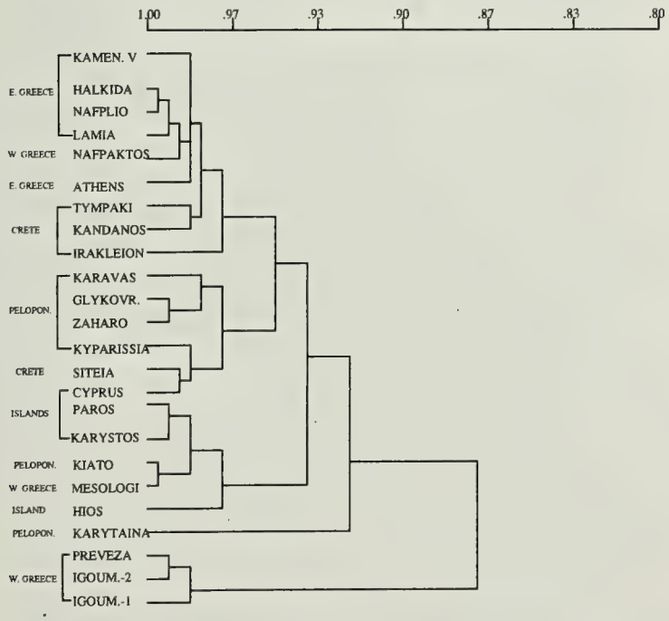


Figure 4. Cluster analysis of complete linkage method using Nei's genetic identity between populations of *Helix aspersa*.

This result is in accordance with the findings of Mazon *et al.* (1988) in *C. nemoralis*.

The dendrogram found by cluster analysis of complete linkage method using Nei's identity (figure 4) showed that there was not a considerable distance among the populations studied, apart from the group of northwestern Greek populations (I = 0.88). This slight differentiation may be due to their geographical distance or to

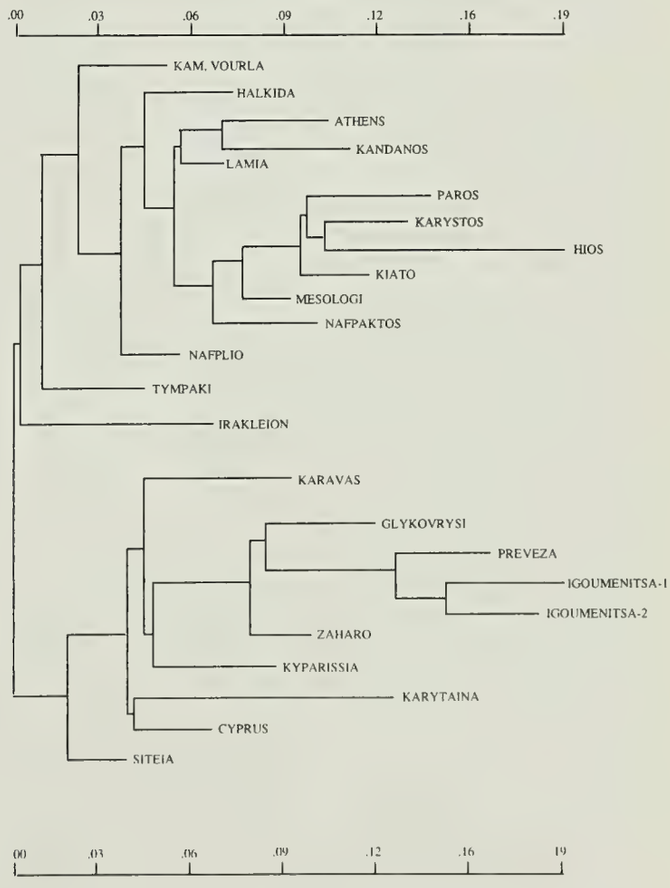


Figure 5. Distance—WAGNER's tree based on allelic frequencies of 24 populations of *Helix aspersa*.

Table 7. Average genetic distance within regions.

Regions	No. of population	1	2	3	4	5	6
1. N. Western Greece	3	0.010	—	—	—	—	—
2. Central Greece	7	0.084	0.010	—	—	—	—
3. Peloponessos	7	0.071	0.029	0.032	—	—	—
4. Aegean Islands	2	0.093	0.024	0.037	0.020	—	—
5. Crete	4	0.088	0.027	0.031	0.037	0.016	—
6. Cyprus	1	0.079	0.022	0.020	0.031	0.023	—

climatic reasons, if we take into account that two of their alleles whose frequencies were high in that region (GPD-1B, LAP-1D), were correlated positively with mean annual precipitation. Such correlations might arise simply by chance in such a restricted geographical survey, and a larger number of independent correlations with the same climatic variable is needed to test a real association. However, climatic selection, at least in some cases, may be involved in the genetic divergence of populations.

Since the island populations were not characterized by a lower degree of heterozygosity and a higher degree of genetic distance from the continental populations, stochastic forces (*e.g.*, founder effect followed to genetic drift) were not the only ones to play an important role in the genetic structuring. Rather, it seems that a combination of forces is responsible for genetic differentiation. Being edible, this species has been introduced by man to many islands in a random way; its populations on an island may either originate from a small sample coming from any region, or may have evolved from the introgression of the introduced and the preexisting populations.

The genetic distance found between the populations of the regions from northwestern Greece and the rest of the populations was not as high as that found by Madec (1991) between French and Algerian populations of *Helix aspersa*. On the basis of our results we cannot support the presence of two geographical races. The northwestern Greek populations comprised a separate group (figure 5), but populations from Crete and Peloponesos, which we expected to be genetically different, were not.

ACKNOWLEDGMENTS

We thank S. E. R. Bailey (University of Manchester, UK) for reading this paper. Thanks are also due to D. Sioula, E. Alpogianni and N. Kifonidis for their technical help. Financial support was provided by the Minister of Research and Technology in Greece.

LITERATURE CITED

Albuquerque de Matos, R. M. 1984. Genetics of shell ground colour in *Helix aspersa* I. Colour locus, uniform and their interactions. Heredity 53(1):11–20.

Albuquerque de Matos, R. M. 1989. Contribution of genetics to snail farming and conservation. *In:* BCPC monograph no. 41. Slugs and snails in world agriculture, 11–18.

Ashton, G., C. Braden, and A. Braden. 1961. Serum, -globuline polymorphism in mice. Australian Journal of Experimental Biology and Medical Sciences 14:248.

Bar, Z. 1978. Variation and natural selection in shell thickness of *Theba pisana* along climatic gradients in Israel. Journal of Molluscan Studies 44:322–326.

Baur, B. 1984. Dispersion, Bestandesdichte und Diffusion bei *Arianta arbustorum* (L.) (Mollusca, Pulmonata). Ph.D. thesis, Univ. of Zurich.

Bleakney, M. M., C. C. Fleming, and R. J. Marks. 1989. Genetic and phenotypic variation in allopatric populations of *Helix aspersa* (Müller): a preliminary report. BCPC Mono. no. 41. Slugs and snails in world agriculture, 319–326.

Chevalier, H. 1977. La variabilité de l'escargot Petit-gris *Helix aspersa* Müller. Bulletin Muséum National Histoire Naturelle. 3e série, 448(311):425–442.

Cook, L. M. 1965. Inheritance of shell size in the snail *Arianta arbustorum*. Genetics 53:415–425.

Crook, S. J. 1982. Genetic studies of allozyme variation in leucine aminopeptidase in the land snail *Helix aspersa* (Müller). Journal of Molluscan Studies 48:362–365.

Goodfriend, G. A. 1986. Variation in land snail shell form and size and its causes: a review. Systematic Zoology 35(2): 204–223.

Haralambidis, S., M. Papazachariadou, M. Lazaridou-Dimitriadou, C. Himonas, and M. Kattoulas. 1985. Contribution to the parasitological study of the commercialized snails of Greece. I. Parasites of the snails *Helix aspersa* Müller and *Helix lucorum* L. from the Greek mainland. Biologia Gallo-Hellenica X:143–148.

Johnson, M. S. 1976. Allozymes and area-effects in *Cepaea nemoralis* on the western Berkshire downs. Heredity 36: 105–121.

Lazaridou-Dimitriadou, M., M. Kattoulas, and A. Staikou. 1983. Searching for the factors that provoke differences in the size and weight of snails (*Helix aspersa* Müller) from two different populations, one from the island of Krete and the other from Peloponesos (Greece). Journal of Molluscan Studies, Supplement 12A:89–93.

Madec, L. 1989. Origine et importance des differances affectant la forme et la taille des oeufs chez l'escargot Petit-gris *Helix aspersa* Müller. Haliotis, 19:143–152.

Madec, L. 1991. Variabilité enzymatique dans quelques populations naturelles de l'espèce *Helix aspersa* Müller. Comptes Rendus de l'Academie des Sciences, Paris, 313, Série III: 339–344.

Madec, L. and J. Duguzan. 1987. Etude de quelques facteurs affectant la reproductions de l'escargot Petit-gris *Helix aspersa* Müller. Haliotis 16:77–91.

Mahalanobis, P. C. 1936. On the generalised distance in statistics. Proceedings of the National Institute of Sciences of India 2:49–55.

Mazon, L. I., A. Vicario, M. A. M. de Pancorbo, A. I. Aguirre, A. Estomba, and C. M. Lostao. 1988. North/south differentiation in the distribution of *Cepaea nemoralis* in Spain. Heredity 61:189–197.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590.

Sacchi, C. F. 1958. Les Mollusques terrestres dans le cadre des relations biogéographiques entre l'Afrique du Nord et l'Italie. Vie et Milieu 9(1):11–52.

Selander, R. K. and D. W. Kaufman. 1975. Genetic structure of the populations of the brown snail (*Helix aspersa*) I. Macrogeographic ratiation. Evolution 29:385–401.

Selander, R. K. and H. Ochman. 1983. Genetics and evolution. *In:* Isozymes: current topics in biological and medical research, Vol. 10. Alan R. Liss, Inc., New York, p. 93–123.

Smith, I. 1976. Chromatographic and electrophoretic techniques, Vol. 2, 2nd ed. Willliam Heinemann Medical Books C.T.T., London.

Sneath, P. and R. Sokal. 1963. Principles of numerical taxonomy. W. H. Freeman, San Francisco.

Swofford, D. L. and R. B. Selander. 1981. BIOSYS-1: a FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. Journal of Heredity 72:281–283.

White, M. J. D. 1978. Modes of speciation. W. H. Freeman, San Francisco.

Wright, S. 1978. Evolution and the genetics of populations, Vol. 4. Variability within and among natural populations. University of Chicago Press, Chicago.

Application of Isoelectric Focusing in Molluscan Systematics

Thierry Backeljau[1]
Karin Breugelmans[2]
Herwig Leirs[2]
Teresa Rodriguez[3]
Dimitri Sherbakov[4]
[1] Royal Belgian Institute of Natural Sciences
Vautierstraat 29
B–1040 Brussels, Belgium
[2] University of Antwerp (RUCA)
Department of Biology Evolutionary Biology Group
Groenenborgerlaan 171
B–2020 Antwerp, Belgium

Tatyana Sitnikova[4]
Jean-Marie Timmermans[1]
Jackie L. Van Goethem[1]
Erik Verheyen[1]
[3] Department of Animal Biology
Faculty of Biology
University of Santiago de Compostela
E–15706 Santiago de Compostela, Spain
[4] Limnological Institute
Siberian Division of the Russian Academy of Science
Irkutsk, Russia

ABSTRACT

The application of isoelectric focusing (IEF) in molluscan systematics is reviewed and illustrated using literature data and unpublished analyses. IEF can be used as any other electrophoretic method, but is most appropriate for: (1) generating complex species-specific banding profiles, (2) assessing overall genetic similarities, (3) supplementing conventional electrophoretic techniques by resolving hidden protein variation and (4) investigating minute organisms.

Key Words: Mollusca; systematics; phylogeny; population genetics; protein electrophoresis; isoelectric focusing.

INTRODUCTION

Protein electrophoresis is still one of the most frequently used molecular techniques in systematics and population genetics. The basis for this technique is that mobility differences of proteins in an electric field reflect changes in their amino acid composition and thus mirror differences at the gene level. Hence, it is a simple, indirect, way to look at gene pools.

However, as conventional electrophoretic methods only detect mobility or molecular weight differences, they may fail to resolve hidden protein heterogeneity caused by amino acid replacements that are not accompanied by substantial charge and/or molecular mass alterations (Coyne *et al.*, 1979; Ramshaw *et al.*, 1979; Singh, 1979; Ferguson, 1980).

Other separation methods such as Isoelectric Focusing (IEF), may reduce this problem. IEF separates proteins according to their isoelectric point (pI) (*e.g.*, Righetti, 1983). To this end one creates a pH gradient by electrophoretic segregation of "carrier ampholytes" (*i.e.* synthetic polyaminopolycarboxylic acids) in a supporting medium. Proteins placed in such a pH gradient will move according to their net charge until they reach a point where the pH equals their pI so that their net charge becomes zero and no further migration occurs. In this way, IEF can separate protein fractions with pI values differing by only 0.01 pH units (Drysdale, 1975; Righetti, 1983). Such resolution by charge is not normally obtainable by other electrophoretic methods and IEF is therefore well suited to examine hidden heterogeneity (Drysdale, 1975; Ross, 1977; Righetti, 1983; Cicchetti *et al.*, 1990).

In this paper we review the use of IEF in molluscan systematics. We therefore provide a survey of IEF applications, after which we focus on IEF data treatment insofar as this differs from other electrophoretic techniques. For general technical accounts on the method we refer to Righetti (1983) and Whitmore (1990a), even though we present some basic guidelines in Appendix 1. Authorships of the molluscan taxa mentioned are provided in Appendix 2.

REVIEW OF IEF APPLICATIONS IN MOLLUSCAN SYSTEMATICS

Gastropoda: Pulmonata

The first applications of IEF in molluscan systematics we could trace, were published by Saladin *et al.* (1976) who used general egg proteins to distinguish between the bulinids *Bulinus liratus* and *B. obtusispira* from Madagascar. Ross (1977) studied glucose phosphate isomerase patterns in *Bulinus* spp. but drew no conclusions. Subsequently, Rollinson and Southgate (1979) investigated five enzymes in 38 populations of four *B. africanus* group

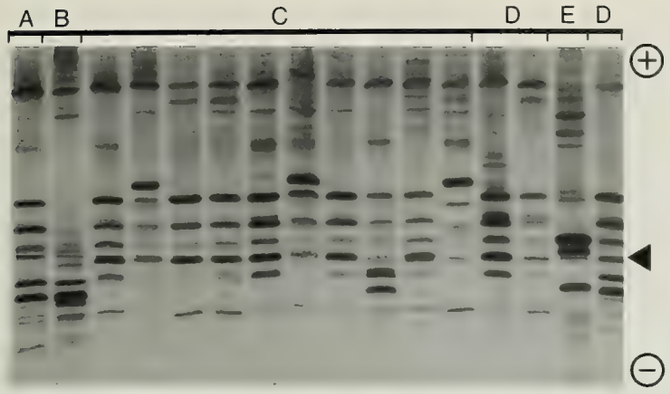


Figure 1. Agarose IEF (pH 4–6.5) of digestive gland esterases in the *Arion hortensis* complex. A-D: *A. distinctus* (A: Brasschaat; B: Wilrijk; C: Hoogstraten; D: Wilrijk). E: *A. hortensis* (Wilrijk).

species in Tanzania and found that *B. nasutus* is clearly differentiated from the other three species. In a more extensive survey of eight of the ten nominal species in the *B. africanus* group, Wright and Rollinson (1979) noted that certain enzyme profile combinations appeared to be associated with some taxa and others with regional distributions. Wright *et al.* (1979) found little heterogeneity within and between populations of *B. senegalensis* (based on five enzymes), but snails parasitized by different trematodes were easily distinguished. Wright and Rollinson (1981) investigated the same five enzymes in 103 populations of the *B. tropicus-truncatus* complex and found that diploid and tetraploid populations were clearly different. These observations were used by Brown and Rollinson (1982) and Brown *et al.* (1982) to characterize *B. truncatus* in the southern part of its distribution and to show that *B. coulboisi* from Lake Tanganyika is only a southern form of *B. truncatus*. Similarly, Brown *et al.* (1986) used IEF enzyme profiles to show that *B. guernei* from West Africa is conpecific with *B. truncatus*, while Southgate *et al.* (1985, 1989) relied on Wright and Rollinson's (1979, 1981) work to demonstrate that diploid Kenyan populations of *B. tropicus* can transmit the fluke *Schistosoma bovis* and that snails parasitized by different trematodes can be separated on the basis of their IEF profiles. Rollinson and Wright (1984) and Rollinson *et al.* (1990) surveyed several enzyme loci in *B. cernicus* from Mauritius. Allele frequencies at these loci showed clear spatial heterogeneities, but were remarkably consistent over a period of six years. Finally, Brown and Shaw (1989) and Brown *et al.* (1991) used IEF of five enzymes to separate Kenyan *B. tropicus*, *B. truncatus* and *B. permembranaceus*.

Backeljau (1985) conducted an IEF analysis of esterases in sibling species of the *Arion hortensis* complex (Figure 1). Mean intra- and interspecific band similarity values showed that *A. hortensis*, *A. distinctus* and *A. owenii* are clearly different. The same study also illustrated the striking difference between monomorphic IEF profiles of uniparental species (*e.g.*, *A. intermedius*) and the highly variable profiles of allogamous species (*e.g.*, *A. hortensis* and *A. distinctus*). Because of this, Backeljau (1985) assumed that *A. owenii* might be a facultative uniparental species. However, the specimens investigated were probably highly inbred for they belonged to a captive stock derived from the original material used by Davies (1977, 1979). Hence the lack of variation in these profiles may have been caused by sustained inbreeding as well. Finally, since the IEF profiles of *A. owenii* were very similar to those of *A. intermedius*, Backeljau (1985) suggested that the latter species belongs to the same subgenus as *A. hortensis* s.l. This conclusion was further elaborated by Backeljau and De Bruyn (1990). In a similar way Backeljau *et al.* (1987) dealt with the *A. fasciatus* complex (Figure 2). IEF profiles of albumen gland proteins and esterases clearly separated three presumed species: *A. fasciatus*, *A. circumscriptus* and *A. silvaticus*. In contrast to *A. hortensis* and *A. distinctus*, but comparable to *A. intermedius*, *A. fasciatus* s.l. revealed a remarkable "intraspecific" profile constancy (Figure 2), even over large geographic distances. This was inter-

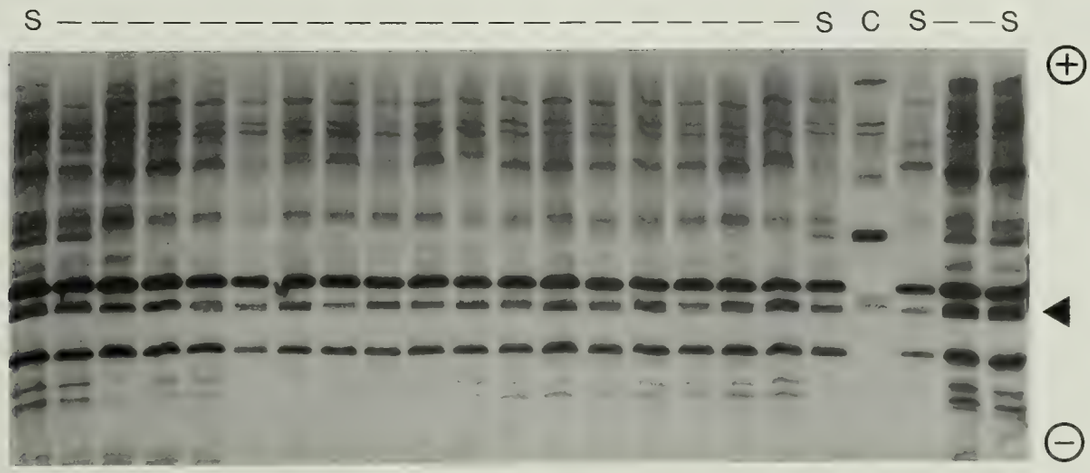


Figure 2. Agarose IEF (pH 4–6.5) of digestive gland esterases in *Arion circumscriptus* (C) and *A. silvaticus* (S).

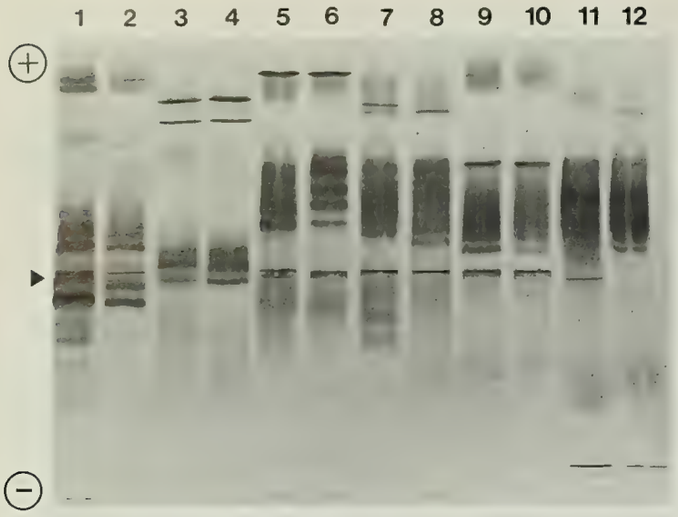


Figure 3. Detection of a cryptic species (genus *Arion*, subgenus *Kobeltia*) by agarose IEF of albumen gland proteins in a 4–6.5 pH gradient. 1–2: *A. (K.) fagophilus* (Alsasua, Spain); 3–4: *A. (K.) intermedius* (3. Boeckhoute, Belgium; 4. Hamburg, Germany); 5–6: *A. (K.) distinctus* (Deurne, Belgium); 7–8: unidentified *A. (K.) hortensis* like species from southern France; 9–10: *A. (K.) hortensis* (Wilmslow, U.K.); 11–12: *A. (K.) owenii* (11. Buncrana, Ireland; 12. London, U.K.).

preted as indicating uniparental reproduction and therefore Backeljau *et al.* (1987) suggested to consider *A. fasciatus* s.l. as an agamospecies complex. A weight analysis suggested that the putative albumen gland protein polymorphism in *A. circumscriptus* was not due to developmental differences. The limited esterase variation, on the contrary, was assumed to be environmentally or physiologically determined (*e.g.*, Oxford, 1975, 1978). Backeljau and De Winter (1987), finally, characterized albumen gland protein profiles of two paratypes of *A. fagophilus* in a qualitative IEF comparison of 10 arionid species. This work revealed a fundamental difference in the albumen gland proteins of the subgenera *Kobeltia* and *Carinarion* on the one hand, and *Arion* and *Mesarion* on the other. A review of the use of albumen gland proteins in arionid systematics was presented by Backeljau (1989). In this context, Figure 3 shows an unpublished comparison of albumen gland profiles of six arionids, indicating that an *Arion (Kobeltia) hortensis*-like slug from southern France differs so much from three morphologically extremely similar species (*A. (K.) hortensis*, *A. (K.) distinctus* and *A. (K.) owenii*), that it probably belongs to another (undescribed?) species.

In order to supplement morphological observations Manga-Gonzalez and Rollinson (1986) surveyed five enzymes to differentiate seven *Helicella* species. Two enzymes, malate dehydrogenase and glucosephosphate isomerase, were sufficient to separate all taxa.

Brito (1992) conducted a preliminary qualitative IEF analysis of esterases in seven species of Zonitidae, representing three genera and three subgenera. This work showed that IEF is also useful for taxonomic purposes in this group.

Gastropoda: Prosobranchia

Using IEF of esterases and general proteins Sella and Badino (1980) demonstrated that Mediterranean *Patella coerulea* and *P. aspera* are distinct, yet closely related, species, while *P. lusitanica* is very different (nomenclature used by Sella & Badino, 1980).

In order to find taxon specific IEF profiles, Viyanant *et al.* (1985) analyzed 12 specific enzymes in two species and one subspecies of *Bithynia* in Thailand. Their work showed that *B. funiculata* and *B. siamensis siamensis* differ consistently in four enzymes, while the subspecies *B. siamensis siamensis* and *B. siamensis goniomphalos* only differ in their esterase profiles.

Unpublished preliminary IEF patterns of esterases and general proteins of *Baicalia* species from Lake Baikal (Russia) illustrate the performance of automated IEF using PhastSystem (see Appendix 1). Figure 4 shows interpopulation esterase heterogeneity in *B. costata*, while Figure 5 compares esterase profiles of *B. costata* and *B. turriformis*. Genetic variation at a monomeric, diallelic esterase locus in this latter species is illustrated in Figure 6, while Figure 7 shows the monomorphic profiles of *B. bithyniopsis*. Finally, two presumed species in the *B. herderiana* complex, *viz. B. ventrosula* and *B. herderiana laevis*, reveal variable general protein profiles, but little or no interspecific differentiation (Figure 8).

Qualitative IEF profiles were also used by Nyumura and Hosokawa (1993) to separate two morphotypes of the apple snail *Pomacea canaliculata* in Japan.

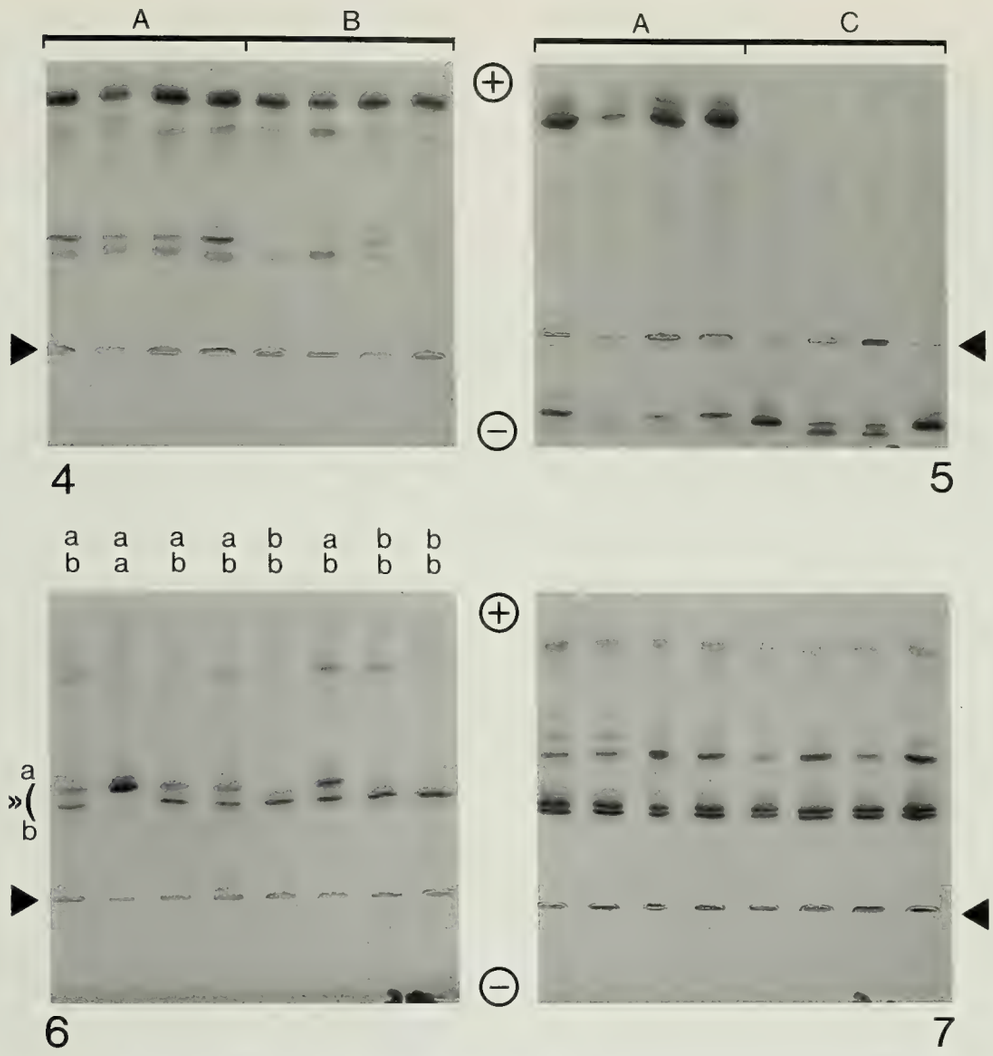
Mill and Grahame (1988) obtained a "reasonable" separation of the morphologically extremely similar *Littorina saxatilis* and *L. arcana* after IEF of non-specific esterases. In addition it was shown that *L. saxatilis* is more variable and heterogeneous than *L. arcana*. Similar results were reported by Dytham *et al.* (1992) and Mill and Grahame (1992), who also observed a clinal change in esterase variation in both periwinkles.

Bivalvia

Günther and Hinz (1986) used IEF of amylases to separate two morphologically similar *Pisidium* species, *viz. P. personatum* and *P. nitidum*. They also noted that two alleles detected by native agarose gel electrophoresis were not resolved by IEF. Yet, subsequently Günther and Hinz (1988) remarked that IEF of amylases was superior to agarose gel electrophoresis in differentiating 15 *Pisidium* and three *Sphaerium* species. The same authors also analyzed phosphoglucomutase with IEF and this, combined with the amylase data, allowed them to confirm: (1) the close relationship between *P. hibernicum* and the group composed of *P. henslowanum*, *P. supinum* and *P. lilljeborgii*, (2) the close relationship between *P. pulchellum* and *P. subtruncatum* and (3) the separate position of *P. amnicum*.

Cephalopoda

Brahma and Lancieri (1979) assessed phylogenetic relationships between *Octopus vulgaris*, *Sepia officinalis*



Figures 4–7. Performance of PhastSystem in a preliminary IEF analysis of esterases in total body homogenates of some *Baicalia* spp. from Lake Baikal. **4.** Interpopulation heterogeneity in *B. costata* (A: Dva Brata; B: Varnachka) (pH 3–9). **5.** Interspecific differentiation between *B. costata* (A: Dva Brata) and *B. turriformis* (C: Dva Brata) (pH 4–6.5). **6.** Genetic variation at a diallelic monomeric esterase in *B. turriformis* (first four specimens from Dva Brata, next four from Varnachka; genotypes are indicated above each lane) (pH 3–9). **7.** Intrapopulation homogeneity of *B. bithyniopsis* (Bolskije Koty) (pH 3–9).

and *Loligo vulgaris* using IEF and immuno-IEF (= IEF followed by an immunodiffusion test against antisera) of eye lens proteins. Immuno-IEF showed a closer relationship between *Sepia* and *Loligo*, than between either of these two and *Octopus*. "Classical" IEF, on the contrary was uninformative as only *Octopus* yielded interpretable IEF profiles.

Lévy *et al.* (1988) used IEF of general proteins of mantle extracts to separate two sibling species of Brazilian *Eledone*, *viz. E. massyae* and *E. gaucha*.

IEF DATA ANALYSIS

The preceding review shows that IEF data have been used in three ways: (1) qualitatively, by seeking taxon specific banding profiles, (2) phenetically, by calculating band pattern similarities and (3) genetically, by interpreting the profiles in terms of loci, alleles and genotypes.

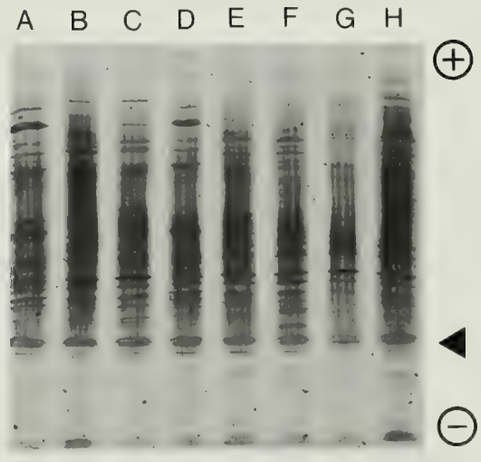
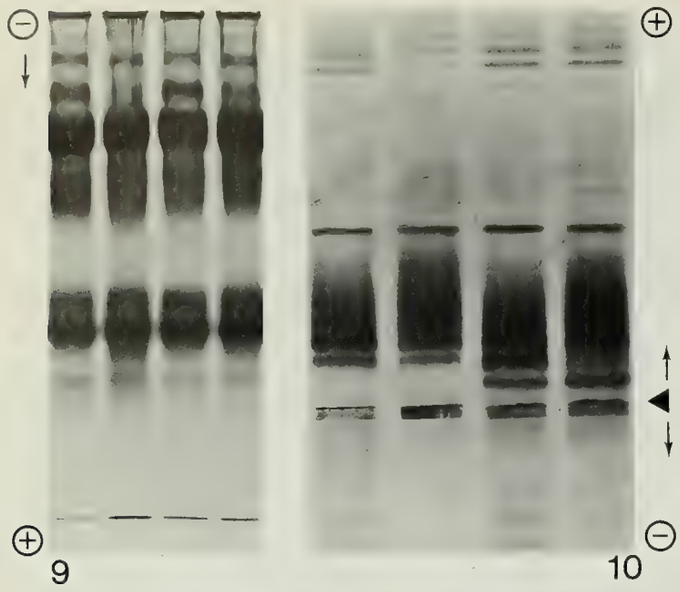


Figure 8. General silver staining of proteins in total body homogenates of *Baicalia ventrosula* (A-D) and *B. herderiana laevis* (E-H) collected at Bolskije Koty (pH 4–6.5).



Figures 9–10. Hidden heterogeneity among albumen gland protein profiles (Coomassie staining) of *Arion hortensis* (Mortsel, Belgium). **9.** Vertical polyacrylamide gel electrophoresis in a 7% gel showing no interindividual variation. **10.** Agarose IEF of the same specimens in a 4–6.5 pH gradient resolving hidden variation in the protein bands near the application site. Migration patterns are indicated by arrows.

Qualitative analyses will not be dealt with further, as they are amply illustrated in our review.

Phenetic analyses are usually performed on banding profiles for which no genetic interpretation is possible (*e.g.*, general protein patterns, uniparental organisms, etc.). Such patterns are compared by the band-counting method (Ferguson, 1980), which treats each band as a distinct character. To this end gels are examined on a light table and adjacent profiles are compared pairwise two under a magnifying lens. The "resemblance" between two profiles can then be expressed by a similarity index, which usually relies on a ratio between shared and unique bands (*e.g.*, Lawson *et al.*, 1980). The simplest index was defined by Ferguson (1980):

$$S_F = \frac{c}{m}$$

where c = number of shared bands and m = maximum number of bands in one of the two compared profiles (*e.g.*, Munuswamy, 1982; Backeljau, 1985; Backeljau *et al.*, 1987; Radice *et al.*, 1988; Verheyen *et al.*, 1991; Phillips *et al.*, 1992).

Three other binary similarity indices have also been used for electrophoretic data. The coefficient of Marczewski and Steinhaus is defined as:

$$S_M = \frac{c}{a_T + b_T - c}$$

where c = number of shared bands, a_T = total number of bands in profile A and b_T = total number of bands in profile B (*e.g.*, Sywula and Bartkowiak, 1978). The matching coefficient of Jaccard (S_J) and its modification (S_{JM}) by Czekanowski [often attributed to Dice or Sörensen (Sneath & Sokal, 1973; Clifford & Stephenson, 1975)] are given by:

$$S_J = \frac{c}{a_U + b_U + c} \qquad S_{JM} = \frac{2c}{a_U + b_U + 2c}$$

where c = number of shared bands, a_U = number of unique bands in profile A and b_U = number of unique bands in profile B. (*e.g.*, Sella & Badino, 1980; Stoddart, 1983; Riutort *et al.*, 1992). S_J has been used by Nixon and Taylor (1977) and Ribas *et al.* (1989) to estimate the time of divergence between noninterbreeding taxa according to Nei's (1971) formula:

$$t = \frac{D}{2cn_T\Omega_a}$$

where $D = -\log_e S_J$, c = the proportion of amino acid substitutions detectable by electrophoresis, n_T = the total number of codons needed to code for a protein and Ω_a = the rate of amino acid substitutions per site per year. However, some assumptions and estimations made by Nei (1971) may not be applicable to IEF profiles of general proteins, because one cannot assign band homologies. Moreover, S_J makes the unrealistic assumption that each band is a unique protein species. Finally, the estimation of c was based on charge characteristics only and thus needs correction in the light of the resolving power of IEF.

The statistical properties of 39 binary similarity indices (including S_J and S_{JM}) have been compared by Shi (1993), who recommended the use of Jaccard's coefficient (S_J), because this index meets most statistical requirements. S_{JM} performs very well too (Shi, 1993), but gives more weight to shared bands (2c). This may be an undesirable property, since electrophoretic data tend to inflate similarity indices due to the fact that shared bands do not necessarily involve identical proteins (hidden heterogeneity). Even though IEF reduces the likelihood of such chance similarities, it does not eliminate them and therefore S_{JM} may be less appropriate. The statistical properties of S_F and S_M have not yet been investigated. Hence their performance relative to S_J is unknown.

If bands can be characterized unambiguously (*e.g.*, by their pI values), one may construct discrete presence/absence data matrices (see also Nixon & Taylor, 1977), which can be subjected to multivariate ordination methods or parsimony programs (*e.g.*, Thorpe, 1985). This latter approach is conceptually similar to the "independent allele" model, which treats each allele as a distinct character with two states (presence/absence). However, the application of this model is highly questionable if not invalid (*e.g.*, Murphy, 1993) and therefore it seems more appropriate to use Gelfand's similarity as an alternative:

$$S_G = \frac{1}{1 + \Sigma (x_i - y_i)^2}$$

where x_i and y_i are the frequencies of the ith band in

populations X and Y, and Σ is taken over the total number of bands in both populations. Thus contrary to S_F, S_M, S_J and S_{JM}, S_G is not applicable to individual comparisons, but to group comparisons (*e.g.*, Cline *et al.*, 1992).

All similarity indices mentioned can be converted in dissimilarities using:

$$DIS = 1 - S$$

A computer program to calculate S_F, S_M and S_G and the corresponding DIS values has been written by Angus *et al.* (1988). S_J and S_{JM} can be calculated with the program NTSYS-pc (Rohlf, 1993). General accounts on similarity indices and their statistical properties can be found in Constandse-Westermann (1972), Sneath and Sokal (1973) and Clifford and Stephenson (1975).

Similarities or distances can be compared hierarchically. With S_F, S_M, S_J and S_{JM}, for example, three levels of relatedness can be considered: (1) intrapopulational, (2) interpopulational and (3) interspecific (*e.g.*, Backeljau, 1985). Average similarities can then be calculated as the arithmetic means of all values for a given class of comparisons. Differences between these means can be tested with an estimation of the standard error of the difference between two means (Farnsworth, 1978), a Student-t-test or an analysis of variance followed by a Duncan Multiple Range test or a Student-Newman-Keuls test (Sokal & Rohlf, 1981). These statistics require that the data are independent, normally distributed and homoscedastic (Sokal & Rohlf, 1981). Deviations from the latter two assumptions can be dealt with by applying data transformations or nonparametric tests (Sokal & Rohlf, 1981; Hageman, 1992). The statistical treatment of interdependent data (*e.g.*, when single profiles contribute to more than one comparison) is a much more fundamental problem, which also applies to similarity values calculated from other molecular data such as Random Amplified Polymorphic DNA (RAPD) profiles (Chapco *et al.*, 1992). So, if mean S_F, S_M, S_J or S_{JM} values are to be tested as outlined above, one should use each individual in only one comparison, such that a set of independent similarity values is generated. Gelfand's index (S_G), on the contrary, can be compared statistically by resampling techniques such as bootstrapping or jackknifing over bands (*e.g.*, Crowley, 1992), followed by an estimation of variances and confidence intervals using, for example, an approach similar to that of Mueller and Ayala (1982).

Finally, phenetic IEF data can yield information with respect to the overall variability of organisms in relation to environmental characteristics. Mill and Grahame (1988, 1992), for example, expressed esterase band heterogeneity among littorinid populations from different sites and species by calculating the Shannon Wiener diversity index for each sample as:

$$DIVERSITY_{SW} = -\Sigma\, p_i \log_e p_i$$

where p_i is the frequency of the ith band in the sample. More generally, in phenetic protein similarity analyses one must always consider possible environmental, developmental and seasonal variations before taxonomic conclusions may be drawn (*e.g.*, Backeljau *et al.*, 1987).

Next to phenetic analyses, IEF data can also be interpreted genetically (*e.g.*, figure 6; Rollinson & Wright, 1984; Théron *et al.*, 1989; Alstad & Corbin, 1990; Rollinson *et al.*, 1990; Alstad *et al.*, 1991). Yet, such approach is not always possible because IEF may occasionally yield genetically uninterpretable profiles produced by artifactual interactions between carrier ampholytes and proteins (Hare *et al.*, 1978; Righetti, 1983).

As the genetic analysis of IEF data proceeds in exactly the same way as for other electrophoretic data, we refer to the extensive literature on these methods for more details (*e.g.*, Richardson *et al.*, 1986; Nei, 1987; Weir, 1990; Whitmore, 1990a; Hillis & Moritz, 1990). Computer packages and programs for various aspects of electrophoretic data analysis have been published by Swofford and Selander (1981), Suiter *et al.* (1983), Swofford and Berlocher (1987), Farris (1989), Lessios (1990), Weir (1990a, b), Felsenstein (1991), Swofford (1991), Lewis (1992) [see also Whitkus, 1985, 1988], Quesada *et al.* (1992) and Ota (1993). This list is not exhaustive. All these programs were written for PC's and larger computers. Yet, there are also programs for Texas Instruments calculators (Spikell & Blumenberg, 1977; Blumenberg & Spikell, 1978, 1980; Blumenberg, 1981).

DISCUSSION

Because of its generally higher resolving power, IEF provides an effective tool to analyse hidden protein variation not detected by conventional electrophoretic methods (Figures 9–10). It is therefore a complementary technique, which is most conveniently used in conjunction with others. An extreme example of this is two-dimensional (2D) electrophoresis. In this approach proteins are separated by IEF in a first dimension and by, for example, SDS electrophoresis in a second dimension perpendicular to the first one. The resulting profiles often show >100 protein spots and thus provide large data sets. Yet, only very few applications of 2D-electrophoresis in molluscan taxonomy have hitherto been published (*e.g.*, Miyazaky *et al.*, 1988; Tsubokawa & Miyazaki, 1993), but both studies clearly show the utility of this method.

Since IEF concentrates proteins on the basis of their isoelectric points it is also a convenient technique to analyse minute organisms (*e.g.*, Kazmer, 1991). Moreover, single IEF runs combined with general protein stainings, often yield considerably larger numbers of discrete characters (bands) than conventional electrophoretic methods. This may be advantageous when only few specimens can be screened (*e.g.*, rare organisms). As such, IEF also provides a means to perform quick preliminary analyses of particular problems (*e.g.*, in order to plan a more extensive survey using other methods). Finally, IEF seems a most efficient technique for species (taxon) identification, particularly since bands can be identified by their pI values and thus can be compared between gels.

Needless to say that IEF can just as well be used for conventional population genetic applications, even though we believe that other electrophoretic methods will continue to dominate this field because of the lower costs involved.

In conclusion, IEF is a technique that has much to offer, particularly when employed in combination with conventional electrophoresis. Nevertheless, its advantageous features are currently far from fully explored or exploited in systematic malacology.

ACKNOWLEDGEMENTS

We are much indebted to two anonymous referees who provided valuable comments on an earlier draft of this paper. Mr H. Van Paesschen (Brussels) and Mr R. Schaerlaeken (Antwerp) kindly helped with the preparation of the illustrations. This work was supported by F.J.B.R.-grant 2.0004.91 and F.J.B.R.-M.I.-grant 30.35.

LITERATURE CITED

Alstad, D.N. and K. W. Corbin. 1990. Scale insect allozyme differentiation within and between host trees. Evolutionary Ecology 4:43–56.

Alstad, D. N., S. C. Hotchkiss and K. W. Corbin. 1991. Gene flow estimates implicate selection as a cause of scale insect population structure. Evolutionary Ecology 5:88–92.

Angus, R. A., S. W. Hardwick jr. and G.B. Cline. 1988. A computer program that calculates various similarity coefficients between electropherograms. Journal of Heredity 79:74–75.

Backeljau, T. 1985. Estimation of genic similarity within and between *Arion hortensis* s.l. and *A. intermedius* by means of isoelectric focused esterase patterns in hepatopancreas homogenates (Mollusca, Pulmonata: Arionidae). Zeitschrift für zoologische Systematik und Evolutionsforschung 23:38–49.

Backeljau, T. 1989. Electrophoresis of albumen gland proteins as a tool to elucidate taxonomic problems in the genus *Arion* (Gastropoda, Pulmonata). Journal of Medical and Applied Malacology 1:29–41.

Backeljau, T. and L. De Bruyn. 1990. On the infrageneric systematics of the genus *Arion* Férussac, 1819 (Mollusca, Pulmonata). Bulletin van het Koninklijk Belgisch Instituut voor Natuurwetenschappen, Biologie 60:35–68.

Backeljau, T. and A. J. De Winter. 1987. An electrophoretic characterisation of three paratypes of *Arion fagophilus* De Winter, 1986, with notes on the subgeneric division of the genus *Arion* Férussac, 1819 (Mollusca, Pulmonata). Zeitschrift für zoologische Systematik und Evolutionsforschung 25:169–180.

Backeljau, T., S. Z. Ahmadyar, M. Selens, J. Van Rompaey and W. Verheyen. 1987. Comparative electrophoretic analyses of three European *Carinarion* species (Mollusca, Pulmonata, Arionidae). Zoologica Scripta 16:209–222.

Blumenberg, B. 1981. Calculator programs in general genetics IV. The genetic distance indices G_S^2, ϕ_S, G, D_H, and D_R. Journal of Heredity 72:460–462.

Blumenberg, B. and M. Spikell. 1978. Calculator programs in general genetics II. Nei's indices of genetic distance, protein identity, migration rate, and divergence time. Journal of Heredity 69:278–280.

Blumenberg, B. and M. Spikell. 1980. Calculator programs in general genetics III. Latter's indices of heterozygosity, population differentiation, and genetic distance. Journal of Heredity 71:293–294.

Brahma, S. K. and M. Lancieri. 1979. Isofocusing and immunological investigations on cephalopod lens proteins. Experimental Eye Research 29:671–678.

Brito, C. P. D. 1992. Electrophoretic results of a biochemical systematic survey of *Oxychilus (Drouetia) atlanticus* and some other Zonitidae (Gastropoda: Zonitidae). Biological Journal of the Linnean Society 46:145–151.

Brown, D. S. and D. Rollinson. 1982. The southern distribution of the freshwater snail *Bulinus truncatus*. South African Journal of Science 78:290–293.

Brown, D. S. and K. M. Shaw. 1989. Freshwater snails of the *Bulinus truncatus/tropicus* complex in Kenya: tetraploid species. Journal of Molluscan Studies 55:509–532.

Brown, D. S., D. B. Matovu and D. Rollinson. 1982. *Bulinus coulboisi* of Lake Tanganyika: Assessment of its taxonomic position and role as intermediate host for *S. haematobium*. Journal of Natural History 16:673–687.

Brown, D. S., K. M. Shaw, V. R. Southgate and D. Rollinson. 1986. *Bulinus guernei* (Mollusca: Gastropoda) of West Africa: taxonomic status and role as host for schistosomes. Zoological Journal of the Linnean Society 88:59–90.

Brown, D. S., K. M. Shaw and D. Rollinson. 1991. Freshwater snails of the *Bulinus truncatus/tropicus* complex (Basommatophora: Planorbidae) in Kenya: diploid populations. Journal of Molluscan Studies 57:143–166.

Chapco, W., N. W. Ashton, R. K. B. Martel, N. Antonishyn and W. L. Crosby. 1992. A feasibility study of the use of random amplified polymorphic DNA in the population genetics and systematics of grasshoppers. Genome 35:569–574.

Cicchetti, R., G. Argentin and B. Nicoletti. 1990. Investigations on the PGM (phosphoglucomutase) polymorphism by isoelectric focusing in *Drosophila melanogaster*. Biochemical Genetics 28:247–255.

Clifford, H.T. and W. Stephenson. 1975. An introduction to numerical classification. Academic Press, New York, 229 pp.

Cline, G. B., C. W. Hardwick jr. and R. A. Angus. 1992. Diagnostic isozymic markers in stone crabs (Genus *Menippe*) from Mississippi Sound, Apalachicola Bay, and South Florida. *In*: Bert, T.M. (Ed.). Proceedings of a symposium on stone crab (genus *Menippe*) biology and fisheries. Florida Marine Research Publications, No. 50, St. Petersburg, p. 4–9.

Constandse-Westermann, T.S. 1972. Coefficients of biological distance. Anthropological Publications, Oosterhout, 142 pp.

Coyne, J. A., W. F. Eanes, J. A. M. Ramshaw and R. K. Koehn. 1979. Electrophoretic heterogeneity of α-glycerophosphate dehydrogenase among many species of *Drosophila*. Systematic Zoology 28:164–175.

Crowley, P. H. 1992. Resampling methods for computation-intensive data analysis in ecology and ecolution. Annual Review of Ecology and Systematics 23:405–447.

Davies, S. M. 1977. The *Arion hortensis* complex with notes on *A. intermedius* Normand (Pulmonata: Arionidae). Journal of Conchology 29:173–187.

Davies, S. M. 1979. Segregates of the *Arion hortensis* complex (Pulmonata: Arionidae), with the description of a new species, *Arion owenii*. Journal of Conchology 30:123–127.

Dixon, B. R. and H. P. Arai. 1985. Isoelectric focusing of soluble proteins in the characterization of three species of

Hymenolepis (Cestoda). Canadian Journal of Zoology 63: 1720–1723.

Dixon, B.R. and H.P. Arai. 1989. Differentiation of three species of *Hymenolepis* (Cestoidea) using enzyme isoelectric focusing on thin-layer agarose gels. Canadian Journal of Zoology 67:51–54.

Drysdale, J. W. 1975. Isoelectric focusing in polyacrylamide gel. *In*: Catsimpoolas, N. (Ed.). Methods of protein separation Volume 1. Plenum Press, New York/London, p. 93–126.

Dytham, C., J. Grahame and P. J. Mill. 1992. Esterase variation in *Littorina arcana* and *L. saxatilis* at Robin Hood's Bay, Yorkshire. Proceedings of the Third International Symposium on Littorinid Biology, 39–44.

Farnsworth, M. W. 1978. Genetics. Harper and Row, New York, 626 pp.

Farris, J. S. 1989. Hennig86: A PC-DOS program for phylogenetic analysis. Cladistics 5:163.

Felsenstein, J. 1991. PHYLIP (Phylogeny Inference Package) version 3.4 (unpublished manual). University of Washington, Seattle.

Ferguson, A. 1980. Biochemical systematics and evolution. Blackie and Son Limited, Glasgow, 194 pp.

Günther, A. and W. Hinz. 1986. Elektrophoretische Untersuchung der Amylasen (E.C. 3.2.1.1.) bei *Pisidium personatum* und *Pisidium nitidum* (Sphaeriidae; Bivalvia; Mollusca). Zoologischer Anzeiger 216:95–98.

Günther, A. and W. Hinz. 1988. Vergleichende elektrophoretische Untersuchungen an europäischen Sphaeriidae-Arten (Bivalvia). Zoologischer Anzeiger 220:49–58.

Hageman, S. J. 1992. Alternative methods for dealing with nonnormality and heteroscedasticity in paleontological data. Journal of Paleontology 66:857–867.

Hare, D. L., D. I. Stimpson and J. R. Cann. 1978. Multiple bands produced by interaction of a single macromolecule with carrier ampholytes during isoelectric focusing. Archives of Biochemistry and Biophysics 187:274–275.

Harris, H. and D. A. Hopkinson. 1976. Handbook of enzyme electrophoresis in human genetics. North Holland Publishing Company, Amsterdam/Oxford, unpaged.

Herberts, C., J. de Frescheville, M. Quignon and Y.J. Golvan. 1989. A comparative study of the protein composition of the albumen gland of two populations of *Biomphalaria glabrata* (Say) differing in their susceptibility to *Schistosoma mansoni* (Sambon). Journal of Molluscan Studies 55: 373–377.

Hillis, D. M. and C. Moritz (Eds.). 1990. Molecular systematics. Sinauer Associates, Inc., Sunderland, MA. 588 pp.

Holmes, R. S., B. Cheung and J. L. Vandeberg. 1989. Isoelectric focusing studies of aldehyde dehydrogenases, alcohol dehydrogenases and oxidases from mammalian anterior eye tissues. Comparative Biochemistry and Physiology 93B:271–277.

Kazmer, D. J. 1991. Isoelectric focusing procedures for the analysis of allozyme variation in minute arthropods. Annals of the Entomological Society of America 84:332–339.

Keyvanfar, A., D. Rochu and J. M. Fine. 1988. Comparative study of sturgeon oocyte soluble proteins by isoelectric focusing. Comparative Biochemistry and Physiology 90B: 393–396.

Kilias, H. 1988. Isoenzyme patterns as a tool in taxonomy of chlorococcal algae. Electrophoresis 9:613–617.

Lawson, R., D. H. Davies, J. Casey and S. A. Mohammad. 1980. The interpretation of protein separations on polyacrylamide gels. Laboratory Practice 29:1065–1066.

Lessios, H. A. 1990. A program for calculating Nei's genetic distances and their jackknifed confidence intervals. Journal of Heredity 81:490.

Lévy, J. A., M. Haimovici and M. Conceicao. 1988. Genetic evidence for two species to the genus *Eledone* (Cephalopoda: Octopodidae) in South Brazil. Comparative Biochemistry and Physiology 90B:275–277.

Lewis, P. O. 1992. GeneStat-PC 3.3. Unpublished manual, North Carolina State University, Raleigh.

Manga-Gonzalez, Y. and D. Rollinson. 1986. Preliminary observations on the use of isoelectric focusing of enzymes for the identification of *Helicella* species (Mollusca, Stylommatophora) from Leon province (Spain). Iberus 6:155–165.

Mill, P. J. and J. Grahame. 1988. Esterase variability in the gastropods *Littorina saxatilis* (Olivi) and *L. arcana* Ellis. Journal of Molluscan Studies 54:347–353.

Mill, P. J. and J. Grahame. 1992. Clinal changes in esterase variability in *Littorina saxatilis* (Olivi) and *L. arcana* Hannaford Ellis in Southern Britain. Proceedings of the Third International Symposium on Littorinid Biology, 31–37.

Miyazaki, J.-I., R. Ueshima and T. Hirabayashi. 1988. Application of a two-dimensional electrophoresis method to the systematic study of land snails of subgenus *Luchuphaedusa* from southwestern Japan islands. Biological Bulletin 175:372–377.

Morizot, D. C. and M. E. Schmidt. 1990. Starch gel electrophoresis and histochemical visualization of proteins. *In*: Whitmore, D.H. (Ed.). Electrophoretic and isoelectric focusing techniques in fisheries management. CRC Press, Boca Raton, p. 23–80.

Mork, J. 1990. Eggs and embryos. *In*: Whitmore, D.H. (Ed.). Electrophoretic and isoelectric focusing techniques in fisheries management. CRC Press, Boca Raton, p. 297–311.

Mueller, L. D. and F. J. Ayala. 1982. Estimation and interpretation of genetic distance in empirical studies. Genetical Research, Cambridge 40:127–137.

Munuswamy, N. 1982. Electrophoregram studies of three problematic *Streptocephalus* species (Anostraca). Crustaceana 43:262–264.

Murphy, R. W. 1993. The phylogenetic analysis of allozyme data: invalidity of coding alleles by presence/absence and recommended procedures. Biochemical Systematics and Ecology 21:25–38.

Murphy, R. W., J. W. Sites jr., D. G. Buth and C. H. Haufler. 1990. Proteins 1: isozyme electrophoresis. *In*: Hillis, D. M. and C. Moritz (Eds.). Molecular Systematics. Sinauer Associates, Sunderland, MA, p. 45–126.

Nei, M. 1971. Interspecific gene differences and evolutionary time estimated from electrophoretic data on protein identity. American Naturalist 105:385–398.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York, 512 pp.

Nixon, S. E. and R. J. Taylor. 1977. Large genetic distances associated with little morphological variation in *Polycelis coronata* and *Dugesia tigrina* (Planaria). Systematic Zoology 26:152–164.

Nunamaker, R. A. and C. N. McKinnon. 1989. Electrophoretic analyses of proteins and enzymes in *Culicoides variipennis* (Diptera: Ceratopogonidae). Comparative Biochemistry and Physiology 92B:9–16.

Nyumura, N. and T. Hosokawa. 1993. Morphological and electrophoretic comparisons of two types of *Pomacea canaliculata* in Chiba prefecture. The Chiribotan 24:6–8 [in Japanese].

Olsson, I., U.-B. Axiö-Fredriksson, M. Degerman and B. Olsson

1988a. Fast horizontal electrophoresis. I. Isoelectric focusing and polyacrylamide gel electrophoresis using PhastSystem. Electrophoresis 9:16–22.

Olsson, I., R. Wheeler, C. Johansson, B. Ekström, N. Stafström, R. Bikhabhai and G. Jacobson. 1988b. Fast horizontal electrophoresis. II. Development of fast automated staining procedures using PhastSystem. Electrophoresis 9:22–27.

Ota, T. 1993. DISPAN: genetic distance and phylogenetic analysis, version 1.1. Unpublished manual, The Pennsylvania State University, 328 Mueller Laboratory, University Park, PA 16802.

Oxford, G. S. 1975. Food induced esterase phenocopies in the snail *Cepaea nemoralis*. Heredity 35:361–370.

Oxford, G. S. 1978. The nature and distribution of food-induced esterases in helicid snails. Malacologia 17:331–339.

Payan, L. A. and D. W. Dickson. 1990. Comparison of populations of *Pratylenchus brachyurus* based on isozyme phenotypes. Journal of Nematology 22:538–545.

Phillips, M. S., B. E. Harrower, D. L. Trudgill, M. A. Catley and R. Waugh. 1992. Genetic variation in British populations of *Globodera pallida* as revealed by isozyme and DNA analyses. Nematologica 38:304–319.

Privalov, P. L. 1990. Cold denaturation of proteins. Critical Reviews in Biochemistry and Molecular Biology 25:281–305.

Quesada, H., J. Sanjuan and A. Sanjuan. 1992. GENET2: A Basic program for the analysis of deviations from Hardy-Weinberg law. Journal of Heredity 83:460–461.

Rabilloud, T. 1990. Mechanisms of protein silver staining in polyacrylamide gels: A 10-year synthesis. Electrophoresis 11:785–794.

Radice, A. D., R. D. Riggs and F. H. Huang. 1988. Detection of intraspecific diversity of *Heterodera glycines* using isozyme phenotypes. Journal of Nematology 20:29–39.

Ramshaw, J. A. M., J. A. Coyne, and R. C. Lewontin. 1979. The sensitivity of gel electrophoresis as a detector of genetic variation. Genetics 93:1019–1037.

Ribas, M., M. Riutort and J. Baguna. 1989. Morphological and biochemical variation in populations of *Dugesia (G.) tigrina* (Turbellaria, Tricladida, Paludicola) from the western Mediterranean: biogeographical and taxonomical implications. Journal of Zoology (London) 218:609–626.

Richardson, B. J., P. R. Baverstock and M. Adams. 1986. Allozyme electrophoresis - A handbook for animal systematics and population studies. Academic Press, Sydney, 410 pp.

Righetti, P.G. 1983. Isoelectric focusing: Theory, methodology and applications. Elsevier Biomedical Press, Amsterdam, 386 pp.

Riutort, M., K. G. Field, R. A. Raff and J. Baguna. 1992. Enzyme electrophoresis, 18S rRNA sequences, and levels of phylogenetic resolution among several species of freshwater planarians (Platyhelminthes, Tricladida, Paludicola). Canadian Journal of Zoology 70:1425–1439.

Robinson, M. P. 1989. Isoelectric focusing techniques for the identification of plant-parasitic nematodes. *In*: Loxdale, H. D. and J. den Hollander (Eds.). Electrophoretic studies on agricultural pests. Systematics Association Special Volume No. 39, p. 431–442.

Rohlf, F. J. 1993. NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, version 1.80. Exeter Software, Setauket, New York.

Rollinson, D. and V. R. Southgate. 1979. Enzyme analyses of *Bulinus africanus* group snails (Mollusca: Planorbidae) from Tanzania. Transactions of the Royal Society of Tropical Medicine and Hygiene 73:667–672.

Rollinson, D. and C. A. Wright. 1984. Population studies on *Bulinus cernicus* from Mauritius. Malacologia 25:447–463.

Rollinson, D., R. A. Kane, A. Warlow and V. R. Southgate. 1990. Observations on genetic diversity of *Bulinus cernicus* (Gastropoda: Planorbidae) from Mauritius. Journal of Zoology (London) 222:19–26.

Ross, G. C. 1977. Analysis by isoelectric focusing of phosphoglucoseisomerases in *Schistosoma* species and their snail hosts. Proceedings of the Analytical Division of the Chemical Society 14:76–79.

Saladin, B., A. Degrémont and N. Weiss. 1976. Isoelectric focusing in the taxonomy of Bulinid snails. Acta Tropica 33:376–379.

Sattler, P. W. and L. R. Hilburn. 1985. A program for calculating genetic distance, and its use in determining significant differences in genetic similarity between two groups of populations. Journal of Heredity 76:400.

Sella, G. and G. Badino. 1980. Differenze elettroforetiche tra le specie mediterranee di *Patella* e loro impiego nella tassonomia del genere. Atti IV Congresso S.M.I., Siena, 6–9 Ottobre 1978:357–368.

Sévigny, J.-M. and P. Odense. 1985. Comparison of isoenzyme systems of calanoid copepods by use of ultrathin agarose gel isoelectric focusing techniques. Comparative Biochemistry and Physiology 80B:455–461.

Shi, G.R. 1993. Multivariate data analysis in palaeoecology and palaeobiogeography - a review. Palaeogeography, Palaeoclimatology, Palaeoecology 105:199–234.

Singh, R. S. 1979. Genic heterogeneity within electrophoretic "alleles" and the pattern of variation among loci in *Drosophila pseudoobscura*. Genetics 93:997–1018.

Sneath, P. H. A. and R. R. Sokal. 1973. Numerical taxonomy. W. H. Freeman and Company, San Francisco, 573 pp.

Sokal, R. R. and F. J. Rohlf. 1981. Biometry (2nd edition). W. H. Freeman and Company, New York, 859 pp.

Southgate, V. R., D. S. Brown, D. Rollinson, G. C. Ross and R. J. Knowles. 1985. *Bulinus tropicus* from Central Kenya acting as a host for *Schistosoma bovis*. Zeitschrift für Parasitenkunde 71:61–69.

Southgate, V.R., D.S. Brown, A.Warlow, R.J. Knowles and A. Jones. 1989. The influence of *Calicophoron microbothrium* on the susceptibility of *Bulinus tropicus* to *Schistosoma bovis*. Parasitology Research 75:381–391.

Spikell, M. and B. Blumenberg. 1977. Calculator programs in general genetics I. Computing genetic distance. Journal of Heredity 68:187–190.

Stoddart, J.A. 1983. The accumulation of genetic variation in a parthenogenetic snail. Evolution 37:546–554.

Suiter, K. A., J. F. Wendel and J. S. Case. 1983. LINKAGE–1: a Pascal computer program for the detection and analysis of genetic linkage. Journal of Heredity 74:203–204.

Swofford, D. L. 1991. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0s. Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.

Swofford, D. L. and S. H. Berlocher. 1987. Inferring evolutionary trees from gene frequency data under the principle of maximum parsimony. Systematic Zoology 36:293–325.

Swofford, D. L. and R. B. Selander. 1981. BIOSYS–1: a Fortran program for the comprehensive analysis of electrophoretic data in population genetics and systematics. Journal of Heredity 72:281–283.

Sywula, T. and S. Bartkowiak. 1978. Preliminary study of the isoenzymes of Ostracoda. Crustaceana 35:265–272.

Taylor, D. J., T. L. Finston and P. D. N. Hebert. 1994. The 15% solution for preservation. Trends in Ecology and Evolution 9:230.

Théron, A., P. Brémond and D. Imbert-Establet. 1989. Allelic frequency variations at the MDH-1 locus within *Schistosoma mansoni* strains from Guadeloupe (French West Indies): Ecological interpretation. Comparative Biochemistry and Physiology 93B:33–37.

Thorpe, R. S. 1985. A numerical analysis of isoelectric focused keratin monomers: affinities of some western Indian Ocean green geckos. Biochemical Systematics and Ecology 13: 63–69.

Tsubokawa, R. and J.-I. Miyazaki. 1993. Application of a two-dimensional electrophoresis method to the systematic study of Notaspidea (Mollusca: Opisthobranchia). The Veliger 36:209–214.

Verheyen, E., J. Van Rompaey, R. Blust and W. Decleir. 1991. The eye-lens protein pattern of eight Lake Tanganyika cichlids studied by isoelectric focusing. Annalen van het Koninklijk Museum voor Midden-Afrika, Zoölogische Wetenschappen 262:29–32.

Viyanant, V. and E.S. Upatham. 1985. Isoenzyme analyses of Malayan *Schistosoma*, *S. mekongi* and *S. japonicum* by isoelectric focusing in polyacrylamide gel. Southeast Asian Journal of Tropical Medicine and Public Health 16: 539–545.

Viyanant, V., E. S. Upatham and S. Siriteramongkol. 1985. Enzyme analyses of *Bithynia* (Mollusca: Bithyniidae) by isoelectric focusing. Malacological Review 18:15–20.

Weir, B. S. 1990a. Genetic data analysis. Sinauer Associates, Sunderland, MA, 377 pp.

Weir, B. S. 1990b. Intraspecific differentiation. *In*: Hillis, D. M. and C. Moritz (Eds.). Molecular systematics. Sinauer Associates, Sunderland, MA, p. 373–410.

Westheide, W. and V. Brockmeyer. 1992. Suggestions for an index of enchytraeid species (Oligochaeta) based on general protein patterns. Zeitschrift für zoologische Systematik und Evolutionsforschung 30:89–99.

Whitkus, R. 1985. FORTRAN program for computing genetic statistics from allelic frequency data. Journal of Heredity 76:142.

Whitkus, R. 1988. Modified version of GENESTAT, a program for computing genetic statistics from allelic frequency data. Plant Genetics Newsletter 4:10.

Whitmore, D. H. 1986. Identification of sunfish species by muscle protein isoelectric focusing. Comparative Biochemistry and Physiology 84B:177–180.

Whitmore, D. H. (Ed.). 1990a. Electrophoretic and isoelectric focusing techniques in fisheries management. CRC Press, Boca Raton, 350 pp.

Whitmore, D. H. 1990b. Isoelectric focusing of proteins. *In*: Whitmore, D.H. (Ed.). Electrophoretic and isoelectric focusing techniques in fisheries management. CRC Press, Boca Raton, p. 81–105.

Wright, C. A. and D. Rollinson. 1979. Analysis of enzymes in the *Bulinus africanus* group (Mollusca: Planorbidae) by isoelectric focusing. Journal of Natural History 13:263–273.

Wright, C. A. and D. Rollinson. 1981. Analysis of enzymes in the *Bulinus tropicus/truncatus* complex (Mollusca: Planorbidae). Journal of Natural History 15:873–885.

Wright, C. A., D. Rollinson and P. H. Goll. 1979. Parasites in *Bulinus senegalensis* (Mollusca: Planorbidae) and their detection. Parasitology 79:95–105.

Appendix 1. Basic guidelines for IEF experimentation.

Sample preparation

Either total body homogenates or specific tissues can be used for IEF. Muscle, mantle, gonad and digestive gland extracts are good sources to resolve specific enzymes, whereas albumen glands, eggs and eye lenses (cephalopods) are more convenient for general protein stainings. Samples have to be prepared on ice to prevent heat denaturation of the proteins. Since salts may distort IEF gradients (Sévigny & Odense, 1985; Robinson, 1989; Whitmore, 1990b), tissues are preferably homogenized in distilled water (*e.g.* Sella & Badino, 1980; Manga-Gonzalez & Rollinson, 1986), to which we add 20% (w/v) sucrose (Backeljau, 1985, 1989). Yet, organic (*e.g.*, Brahma & Lancieri, 1979; Viyanant *et al.*, 1985) or dilute inorganic (*e.g.*, Wright & Rollinson, 1979; Günther & Hinz, 1986, 1988; Mill & Grahame, 1988) buffers are tolerated too. Wright & Rollinson (1979) also added 1.0 mM of dithiothreitol, ϵ-aminocaproic acid and ethylenediamine tetraacetic acid (EDTA) as enzyme stabilizers. Still many more extraction solutions are possible (*e.g.*, Dixon & Arai, 1985; Sévigny & Odense, 1985; Kilias, 1988; Keyvanfar *et al.*, 1988; Holmes *et al.*, 1989; Robinson, 1989; Payan & Dickson, 1990; Phillips *et al.*, 1992).

We add 5 μl extraction solution per mg tissue, but other proportions have been used too: Viyanant *et al.* (1985) and Mill and Grahame (1988) homogenized individual snails in respectively 300 μl and 200 μl buffer; Günther and Hinz (1986, 1988) placed single *Pisidium* specimens in 50 μl solution; Wright and Rollinson (1979) and Sella and Badino (1980) used 1:1 proportions, while Brahma and Lancieri (1979) prepared 2% (w/v) homogenates. Tissues may be homogenized with a pestle and mortar (Mill & Grahame, 1988), a mixer (Sella & Badino, 1980; Backeljau, 1989) or a sonicator (Günther & Hinz, 1986, 1988). Viyanant *et al.* (1985) first homogenized snails with a mixer and subsequently sonicated the suspensions three times for 20 sec at 150 W. Homogenates are subsequently centrifuged during 30–45 min at 18000 $\times$ g (= 13000 r.p.m.) to 27000 $\times$ g (= 15000 r.p.m.) (4°C) (Backeljau, 1985, 1989). Following regimes have also been reported: 25 min at 50000 $\times$ g (Wright & Rollinson, 1979), 10 min at 6000 r.p.m. and 5 min at 12000 r.p.m. (Sella & Badino, 1980), 30 min at 12000 r.p.m. (Viyanant *et al.*, 1985) and 4 min at 5000 r.p.m. (Herberts *et al.*, 1989). Brahma and Lancieri (1979) used glass fiber papers (5 $\times$ 10 mm) to absorb 20 μl extract without centrifugation. Mill and Grahame (1988) centrifuged their homogenates during 2.5 min at 4000 r.p.m. and freeze-dried the supernates. These were rehydrated with distilled water when needed.

Supernates can be stored at or below –70°C. Some proteins, however, may denature at these temperatures

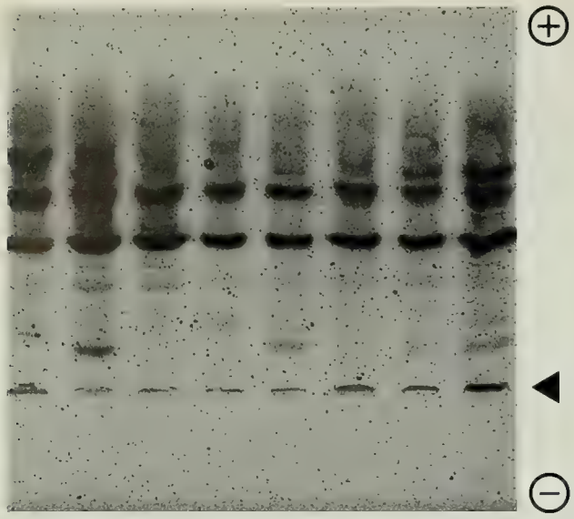


Figure 11. IEF (pH gradient 3–9) patterns of esterases from digestive gland homogenates of *Bukobia* sp. (Mufindi, Tanzania) after more than five years of storage at –70°C. IEF was performed with PhastSystem. Note the granulation caused by undissolved Fast Blue RR (cf. figures 4–7).

(Sévigny & Odense, 1985; Privalov, 1990). For long term storage, it is better to freeze complete individuals or tissues, than to store extracts. Nevertheless, albumen gland homogenates of arionids did not deteriorate over a period of six years, even after repeated freezing and thawing (Backeljau, 1989). Similarly, complete *Bukobia* specimens, stored for more than five years at –70°C, still yielded satisfactory and reproducible esterase profiles (figure 11). The need of fresh material and its storage are disadvantages of protein electrophoresis in general. However, Westheide and Brockmeyer (1992) published protocols for IEF of ethanol-fixed oligochaetes. Taylor *et al.* (1994) reported the possibility of air-drying samples in 15% (w/v) trehalose.

Table 1. Programmed conditions for IEF separations in two pH gradients in mini polyacrylamide gels using PhastSystem (tested with esterases and albumen gland proteins).

		pH 3–9			
Sample appl. down at		2.2			0 Vh
Sample appl. up at		2.3			0 Vh
Extra alarm sound at		2.1			73 Vh
SEP 2.1	2,000 V	2.5 mA	3.5 W	5 °C	75 Vh
SEP 2.2	200 V	2.5 mA	3.5 W	5 °C	15 Vh
SEP 2.3	2,000 V	2.5 mA	3.5 W	5 °C	510 Vh
		pH 4–6.5			
Sample appl. down at		1.2			0 Vh
Sample appl. up at		1.3			0 Vh
Extra alarm sound at		1.1			73 Vh
SEP 1.1	2,000 V	2.0 mA	3.5 W	5 °C	75 Vh
SEP 1.2	200 V	2.0 mA	3.5 W	5 °C	15 Vh
SEP 1.3	1,500 V	4.0 mA	3.5 W	5 °C	510 Vh

Table 2. Programmed silver staining procedure as used with the PhastSystem development unit. EtOH = ethanol, HAc = Acetic acid, TCA = Trichloroacetic acid. Background reducer consists of 2.5 g sodium thiosulphate + 3.7 g Tris in 10 ml reagent grade water; developer consists of 1 ml 2% formaldehyde + 150 ml 2.5% sodium carbonate.

Dev	Solution	In	Out	Time (min)	T (°C)
1	20% TCA	1	0	5	20
2	10% EtOH 5% HAc	3	0	2	50
3	10% EtOH 5% HAc	3	0	4	50
4	5% Glutaraldehyde	4	0	6	50
5	10% EtOH 5% HAc	3	0	3	50
6	10% EtOH 5% HAc	3	0	5	50
7	reagent grade H_2O	5	0	2	50
8	reagent grade H_2O	5	0	2	50
9	0.4% $AgNO_3$	6	0	10	40
10	reagent grade H_2O	5	0	0.5	30
11	reagent grade H_2O	5	0	0.5	30
12	developer	7	0	0.5	30
13	developer	7	0	3.5	30
14	background reducer	8	0	1.5	30
15	reagent grade H_2O	5	0	5	50

Casting and running IEF gels

IEF is usually performed in polyacrylamide (PAA) or agarose gels. Information on PAA gel preparation is provided by Righetti (1983), Viyanant and Upatham (1985), Nunamaker and McKinnon (1989), Robinson (1989), Mork (1990), Whitmore (1990b) and Westheide and Brockmeyer (1992). Agarose recipes can be found in Righetti (1983), Sévigny and Odense (1985), Whitmore (1986), Backeljau (1989) and Dixon and Arai (1989). Most of these references also provide protocols for IEF running conditions. Additional information can be found in Righetti *et al.* (1990) and Whitmore (1990a).

Recently, LKB-Pharmacia introduced an automated electrophoretic unit (PhastSystem) capable of executing, among others, horizontal IEF in mini PAA gels of 50 × 43 × 0.35 mm (Olsson *et al.*, 1988a). In this unit all running conditions are controlled by a programmable microprocessor. It achieves exactly the same resolution as 'manual' IEF in larger gels, but in much shorter times (± 30 min). Table 1 lists our PhastSystem programs for IEF separations of esterases and general proteins in two pH gradients, while figures 4–7 and 11 illustrate some separations obtained with these programs.

Gel staining

After IEF, gels can be stained for either nonspecific proteins or specific enzymes. Recipes for the latter are essentially the same as those published for conventional electrophoresis (*e.g.* Harris & Hopkinson, 1976; Richardson *et al.*, 1986; Morizot & Schmidt, 1990; Murphy *et al.* 1990). Righetti (1983) provided a review of stainings applied in IEF. Some recipes used for molluscs are

given by Wright and Rollinson (1979) and Manga-Gonzalez and Rollinson (1986).

Our recipe for esterase staining is as follows (Backeljau, 1985): dissolve 40 mg Fast Blue RR in a mixture of 25 ml 0.1M KH_2PO_4/NaOH buffer at pH 7.0 (= 6.804 g KH_2PO_4 + 1.164 g NaOH in 1000 ml H_2O), 25 ml H_2O and 2 ml α-naphtylacetate solution (1% w/v α-naphtylacetate in 50% v/v acetone). Before pouring this solution on the gel, it should be filtered to avoid precipitation of undissolved Fast Blue RR (figure 11). Staining takes about 45 min.

General proteins are often stained with Coomassie Brilliant Blue R–250 (= Serva Blue R), as outlined by Backeljau (1989). Silver staining, however, is more sensitive (*e.g.*, Rabilloud, 1990). Several recipes are provided by Righetti (1983). The programmed protocol we follow with PhastSystem (Olsson *et al.* 1988b) is given in table 2.

Agarose and thin PAA gels can be dried and stored after staining. PhastSystem gels can be kept as slides. After prolonged storage (> two years) gels stained for esterases may be covered by a white "dust". This can be washed away by rinsing the gel under gently running tap water. Specific enzyme stainings are less stable for long term storage. Therefore we recommend to photograph or photocopy all gels.

Appendix 2. Systematic list of the molluscan taxa mentioned.

CLASS: GASTROPODA

Subclass: Prosobranchia

Fam. Patellidae
Patella aspera Röding, 1798
Patella coerulea Linnaeus, 1758
Patella lusitanica Gmelin, 1791

Fam. Bithyniidae
Bithynia funiculata Walker, 1927
Bithynia siamensis siamensis Lea, 1856
Bithynia siamensis goniomphalos (Morelet, 1866)

Fam. Baicaliidae
Baicalia (Baicalia) turriformis Dybowski, 1875
Baicalia (Maackia) costata Dybowski, 1875
Baicalia (Eubaicalia) bithyniopsis Lindholm, 1909
Baicalia (Eubaicalia) herderiana laevis Kozhov, 1936
Baicalia (Eubaicalia) ventrosula Lindholm, 1909

Fam. Ampullariidae (= Pilidae)
Pomacea canaliculata (Lamarck, 1804)

Fam. Littorinidae
Littorina saxatilis (Olivi, 1792)
Littorina arcana Hannaford-Ellis, 1978

Subclass: Heterobranchia (partim Pulmonata)

Fam. Planorbidae
Bulinus liratus (Tristram, 1863)
Bulinus obtusispira (Smith, 1882)
Bulinus africanus (Krauss, 1848)
Bulinus nasutus (von Martens, 1879)
Bulinus senegalensis Müller, 1781
Bulinus tropicus (Krauss, 1848)
Bulinus truncatus (Audouin, 1827)
Bulinus coulboisi (Bourguignat, 1888)
Bulinus guernei (Dautzenberg, 1890)
Bulinus cernicus (Morelet, 1867)
Bulinus permembranaceus (Preston, 1912)

Fam. Arionidae
Arion (Kobeltia) hortensis Férussac, 1819
Arion (Kobeltia) distinctus Mabille, 1868
Arion (Kobeltia) owenii Davies, 1979
Arion (Kobeltia) fagophilus de Winter, 1986
Arion (Kobeltia) intermedius Normand, 1852
Arion (Carinarion) fasciatus (Nilsson, 1823)
Arion (Carinarion) circumscriptus Johnston, 1828
Arion (Carinarion) silvaticus Lohmander, 1937

Fam. Urocyclidae
Bukobia sp.

Fam. Helicidae
Helicella sp.

CLASS: BIVALVIA

Fam. Sphaeriidae
Pisidium personatum Malm, 1855
Pisidium nitidum Jenyns, 1832
Pisidium hibernicum Westerlund, 1894
Pisidium henslowanum (Sheppard, 1823)
Pisidium supinum Schmidt, 1851
Pisidium lilljeborgii Clessin, 1886
Pisidium pulchellum Jenyns, 1832
Pisidium subtruncatum Malm, 1855
Pisidium amnicum (Müller, 1774)
Sphaerium sp.

CLASS: CEPHALOPODA

Fam. Sepiidae
Sepia officinalis Linnaeus, 1758

Fam. Loliginidae
Loligo vulgaris Lamarck, 1798

Fam. Octopodidae
Octopus vulgaris Cuvier, 1798
Eledone massyae Voss, 1964
Eledone gaucha Haimovici, 1988

AUTHOR INDEX

SYSTEMATIC INDEX

SUBJECT INDEX

INSTRUCTIONS TO AUTHORS

THE NAUTILUS publishes papers on all aspects of the biology and systematics of mollusks. Manuscripts describing original, unpublished research as well as review articles will be considered. Brief articles, not exceeding 1000 words, will be published as notes and do not require an abstract. Notices of meetings and other items of interest to malacologists will appear in a news and notices section.

Manuscripts: Each original manuscript and accompanying illustrations should be submitted in triplicate. Text must be typed on one side of 8½ × 11 inch white paper, double spaced throughout (including literature cited, tables and figure captions), with at least 1 inch of margin on all sides. All pages must be numbered consecutively. If printed on a word processor, the right margin should be ragged rather than justified. Authors should follow the recommendations of the *Council of Biology Editors Style Manual*, which is available from the Council of Biology Editors, Inc., 9650 Rockville Pike, Bethesda, MD 20814, U.S.A. The first mention of a scientific name in the text should be accompanied by the taxonomic authority, including year. Latin names and words to be printed in italics must be underlined; leave other indications to the editor. Metric and Celsius units are to be used.

The sequence of sections should be: title page, abstract page, introduction, materials and methods, results, discussion, acknowledgements, literature cited, tables, figure captions, figures. The title page should include the title, author's name(s) and address(es). The abstract page should contain the title and abstract, which should summarize in 250 words or less the scope, main results and conclusions of the paper. The abstract may be followed by a maximum of 8 key words. All references cited in the text must appear in the literature cited section and vice versa. In the literature cited section, all authors must be fully identified and listed alphabetically. Follow a recent issue of THE NAUTILUS for bibliographic style, noting that journal titles must be unabbreviated. Information on plates and figures should be cited only if not included in the pagination. Tables must be numbered and each placed on a separate sheet. A brief legend must accompany each table. Captions for each group of illustrations should be typed on a separate sheet and include a key to all lettered labeling appearing in that group of illustrations.

All line drawings must be in black, high quality ink, clearly detailed and completely labeled. Photographs must be on glossy, high contrast paper. All figures are to be consecutively numbered (figs. 1, 2, 3, . . . , NOT figs. 1a. 1b, 1c, . . . NOR plate 1, fig. 1 . . .). Illustrations must be arranged in proportions that will conform with the width of a page (6¾ inches or 171 mm) or a column (3¼ inches or 82 mm). The maximum size of a printed figure is 6¾ by 9 inches or 171 by 228 mm. All illustrations must be fully cropped, mounted on a firm, white backing, numbered, labeled and camera ready. The author's name, paper title and figure number(s) should appear on the back. Original illustrations must be between one and two times the desired final size. It is the author's responsibility that the line weight and lettering are appropriate for the desired reduction. Original illustrations will be returned to the author if requested. Color illustrations can be included at extra cost to the author.

Voucher Material: Deposition of type material in a recognized public museum is a requirement for publication of papers in which new species are described. Deposition of representative voucher specimens in such institutions is strongly encouraged for all other types of research papers.

Processing of Manuscripts: Upon receipt, every manuscript is acknowledged and sent for critical review by at least two referees. These reviews serve as the basis for acceptance or rejection. Accepted manuscripts are returned to the author for consideration of the reviewers' comments. A finalized version of the manuscript is returned to the editor and sent to press. Two sets of proofs are sent to the author for correction. Changes other than typesetting errors will be charged to the author at cost. One set of corrected proofs should be sent to the editor as soon as possible. Authors with institutional, grant or other research support will be billed for page charges at the rate of $60.00 per printed page.

An order form for reprints will accompany the proofs. Reprints may be ordered through the editor.

Manuscripts, corrected proofs and correspondence regarding editorial matters should be sent to: Dr. M.G. Harasewych, Editor, Division of Mollusks, NHB stop 118, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA.


SMITHSONIAN INSTITUTION LIBRARIES